EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

DIRECCIÓN GENERAL DE EDUCACIÓN TECNOLÓGICA INDUSTRIAL Y DE SERVICIOS

Probabilidad y Estadística

Segunda Edición

Aprendizajes Esenciales
febrero-julio de 2021

Academia Nacional de Matemáticas

México 2021 Año de la Independencia

Manual del alumno

In memoriam

Un maestro afecta la eternidad; solo él puede decir donde para su influencia.

Henry Adams

En homenaje a nuestros queridos amigos

Armando Rosas Zepeda
Justino Maza Román
Nelson Gutiérrez Valdés
Ramón Figueroa Saucedo

Índice

Índice 2
Encuadre 8
Propósito 8
Marco teórico 8
Marco referencial 9
Características del curso 9
Recomendaciones para la impartición del curso 11
Competencias a desarrollar en el curso 12
Introducción 13
Justificación 14
Bloque 1 | Manejo de la información 15
1.1 Definición de Estadística 15
1.1.1 Campo de estudio. 15
Actividades de Apertura 15
Actividades de Desarrollo 16
1.1.2 Antecedentes históricos. 17
Actividades de cierre 19
1.2 Población, muestra y tipos de variables. 19
Introducción 19
Actividades de Apertura 20
Actividades de Desarrollo 20
1.2.1 Población 20
1.2.2 Muestra 20
1.2.3 Tipos de variables 21
Actividades de cierre 23
1.3. Niveles de medición de variables 24
Actividades de Apertura 24
Actividades de Desarrollo 24
1.3.1. Nominal 24
1.3.2. Ordinal 24
1.3.3. De intervalo 24
1.3.4. De razón 25
Actividades de cierre 26

1.4 Frecuencias 27
Introducción 27
Actividades de Apertura 27
1.4.1. Frecuencia Absoluta (acumulada) 27
1.4.2. Frecuencia Relativa (acumulada) 27
Actividades de Desarrollo 28
Actividades de cierre 29
Actividades de contexto o Transversales 30
1.5 Tablas de Distribución de Frecuencias 32
1.5.1 Elementos 32
Introducción 32
Actividades de Apertura 32
Actividades de Desarrollo 34
Actividades de cierre 43
Ejercicios Adicionales 49
1.6 Representación gráfica e interpretación. 50
Introducción 50
Actividades de Apertura 50
Actividades de Desarrollo 50
1.6.1. Histograma. 50
1.6.2. Polígono de Frecuencias 53
1.6.3. Ojiva. 54
1.6.4. Gráfico Circular, de sectores o Pastel 55
Actividades de cierre 58
Ejercicios Adicionales 58
Bloque 2 | Medidas de tendencia central 60
2.1 Media 60
Introducción 60
2.1.1. Media aritmética 60
2.2 Mediana y Moda 62
2.2.1 Mediana 62
2.2.2 Moda 64
Actividades de Apertura 67
Actividades de Desarrollo 68
Actividades de cierre 68
Ejercicios Adicionales 69

2.3 Cuartiles, Deciles y Percentiles....71
Introducción 71
2.3.1. Cuartiles....71
2.3.2 Deciles 72
2.3.3 Percentiles 72
Actividades de Apertura 75
Actividades de Desarrollo 76
Actividades de cierre.... 76
Actividades de contexto o Transversales 77
Ejercicios Adicionales 78
Bloque 3 | Medidas de dispersión 80
3.1 Rango 80
Introducción 80
Actividades de Apertura 82
Actividades de Desarrollo 82
Actividades de cierre.... 83
Actividades de contexto o Transversales 84
Ejercicios Adicionales 85
3.2 Desviación media 85
Introducción 85
Actividades de Apertura 86
Actividades de Desarrollo 86
Actividades de cierre.... 87
Actividades de contexto o Transversales 88
Ejercicios Adicionales 88
3.3 Varianza y desviación típica o estándar 89
Introducción 89
3.3.1 Varianza....89
3.3.2 Desviación típica o estándar 90
Actividades de Apertura 90
Actividades de Desarrollo 91
Actividades de cierre.... 92
Actividades de contexto o Transversales 92
Ejercicios Adicionales 93
Bloque 4 | Medidas de forma 95
4.1 Sesgo....95

Introducción 95
Actividades de Desarrollo 95
4.2 Apuntamiento o curtosis 96
Actividades de cierre 97
Bloque 5 | Medidas de correlación 98
5.1 Coeficiente de correlación 98
Introducción 98
Actividades de Apertura 98
Actividades de Desarrollo 98
Actividades de cierre 101
Ejercicios Adicionales 102
5.2 Recta de regresión 105
Introducción 105
Actividades de Apertura 105
Actividades de Desarrollo 105
Actividades de cierre 107
Ejercicios Adicionales 107
5.3 Error estándar de estimación 111
Introducción 111
Actividades de Apertura 111
Actividades de Desarrollo 111
Actividades de cierre 112
Ejercicios Adicionales 112
Bloque 6 | Teoría de conjuntos 114
6.1 Elementos básicos 114
6.1.1 Diagrama de Venn-Euler 114
Introducción 114
Actividades de Apertura 114
Actividades de Desarrollo 114
Actividades de cierre 115
6.1.2 Conjunto 115
Introducción 115
Actividades de Apertura 116
Actividades de Desarrollo 116
Actividades de cierre 117
6.1.3 Subconjunto 118

Introducción 118
Actividades de Apertura 118
Actividades de Desarrollo 118
Actividades de cierre 119
6.1.4 Pertenencia 120
Introducción 120
Actividades de Apertura 120
Actividades de Desarrollo 121
Actividades de cierre 121
6.1.5 Conjunto universo 122
Introducción 122
Actividades de Apertura 122
Actividades de cierre 122
Introducción 123
Actividades de Apertura 123
Actividades de Desarrollo 123
Actividades de cierre 123
6.2 Operaciones con conjuntos 124
6.2.1 Unión 124
Introducción 124
Actividades de Apertura 124
Actividades de Desarrollo 124
Actividades de cierre 125
6.2.2 Intersección 126
Introducción 126
Actividades de Apertura 126
Actividades de Desarrollo 126
Actividades de cierre 127
6.2.3 Complemento 128
Introducción 128
Actividades de Apertura 128
Actividades de cierre 129
6.2.4 Diferencia 130
Introducción 130
Actividades de Apertura 130
Actividades de Desarrollo 130
Actividades de cierre 131

6.3 Representación de operaciones mixtas con hasta tres conjuntos 132
Introducción 132
Actividades de Apertura 132
Actividades de Desarrollo 132
Actividades de cierre 134
Bloque 7 | Técnicas de Conteo 135
7.1 Elementos básicos 135
Introducción 135
Actividades de Apertura 135
Actividades de Desarrollo 136
Actividades de cierre 137
7.1.2. Espacio muestral, evento y diagrama de árbol 137
Introducción 137
7.2 Principio fundamental de la multiplicación 139
Introducción 139
7.3 Principio fundamental de la suma 142
7.3.1 Notación factorial 144
Actividades de Apertura 144
7.4 Combinaciones y Permutaciones 145
Actividades de Desarrollo 145
Actividades de cierre 146
7.4.1 Permutaciones 147
Actividades de Apertura 147
Actividades de Desarrollo 147
Actividades de cierre 148
7.5 Teorema de Bayes 149
Introducción 149
Actividades de Apertura 150
Actividades de cierre 151
Fuentes consultadas 153
Directorio 154
Academia Nacional de Matemáticas 155

Encuadre

Propósito

Desarrollar las competencias necesarias para el aprendizaje de las Matemáticas en los estudiantes de Bachillerato Tecnológico, en los planteles de la DGETI de la República Mexicana, y que le permita lograr el perfil de egreso que exigen los nuevos tiempos, enfrentando la contingencia actual por el SARS-CoV-2 en su permanencia en casa. Cada manual está diseñado, principalmente, para los alumnos con falta de recursos y/o conectividad para que puedan seguir con sus clases desde casa, así mismo, es muy práctico para trabajar con los alumnos que si cuentan con los recursos para llevar sus clases en línea, y como apoyo al docente titular de las asignaturas para propiciar en el alumno, aún en la distancia, el interés de dirigir su automotivación hacia el aprendizaje autodidacta de los contenidos de los programas de estudio vigentes de las asignaturas de Matemáticas de bachillerato en el plan nacional educativo, a través de la construcción de su propio conocimiento y la aplicación pertinente de ellos en su contexto personal y su vida cotidiana desde una óptica crítico-analítica del pensamiento individual.

Marco teórico

Los seres humanos somos capaces de conocer el mundo a través del lenguaje, del análisis lógico-matemático, de la representación espacial, del pensamiento musical, del uso del cuerpo para resolver problemas o hacer cosas, de la propia interpretación del universo, a interrelación con los demás individuos y de una auto comprensión de nosotros mismos. Donde los individuos se diferencian en el nivel e intensidad de sus habilidades y en las formas en que recurre a esas mismas y se les combina para llevar a cabo diferentes labores, para solucionar diversos problemas y progresar en distintos ámbitos.

Las personas aprenden, representan y utilizan el saber de muchos y diferentes modos, estas diferencias desafían al sistema educativo, que hoy en día lucha por contraponerse a las ideas erróneas de que todo el mundo puede aprender los mismos conocimientos, las mismas disciplinas y del mismo modo y que basta con una medida uniforme y universal para poner a prueba el aprendizaje de los alumnos.

Los procesos de aprendizaje de las matemáticas requieren de estrategias que permitan al alumno que las competencias que son adquiridas en la escuela se sitúen en un ambiente cotidiano para relacionar, interpretar inferir y aplicar los saberes a la resolución de problemas.

El desarrollo de habilidades, destrezas y actitudes se relaciona directamente con las condiciones que se deben dar para lograr que los aprendizajes en el estudiante sean significativos y lo más funcional posible.

El proceso de evaluación de las competencias consiste en utilizar los medios que permitan a los alumnos reconocer si los esquemas de actuación aprendidos le son de utilidad, a tal grado que le sirvan para intervenir correctamente ante una situación problemática planteada en la cotidianidad.

Marco referencial

Al analizar los procesos de aprendizaje de las matemáticas, es posible percatarse que los alumnos han experimentado una serie de estrategias por parte de los docentes para que las competencias las transfieran en situaciones de la vida real. Esto exige relacionar, interpretar, inferir, interpolar, inventar, y aplicar los saberes a la resolución de problemas, mediante la intervención en la realidad reflexionando y actuando sobre la acción y reaccionando con responsabilidad ante situaciones imprevistas o contingentes.

El aprendizaje por competencias está directamente relacionado con las condiciones que deben darse para que los aprendizajes sean los más significativos, situados y funcionales posibles.

La evaluación del aprendizaje de competencias responde a la evaluación de contenidos; pero no toda la evaluación está referida a ello. Si consideramos que la evaluación es un aspecto complejo donde convergen diferentes dimensiones, entonces debemos considerar que están implicados procesos de evaluación también complejos.

El proceso de evaluación de las competencias consistirá en utilizar los medios que permitan reconocer si los esquemas de actuación emprendidos por el estudiante pueden serle de utilidad para superar situaciones reales en contextos concretos lo más aproximados a la realidad; para evaluarla es necesario tener datos fiables sobre el grado de aprendizaje de cada estudiante con relación a la competencia implicada, para ello se requiere el uso de instrumentos y medios diversos en función de las características propias de cada competencia y los distintos contextos donde ésta debe o puede llevarse a cabo.

Dado que las competencias están constituidas por uno o más contenidos de aprendizaje, es necesario identificar los indicadores de logro para cada uno de ellos, pero integrados o que se puedan integrar en la competencia correspondiente y el medio para conocer el grado de su aprendizaje será la intervención del estudiante ante la situación problemática planteada. La evaluación bajo el enfoque de competencias no solo implica evaluar el resultado del aprendizaje del alumno, también el proceso de enseñanza-aprendizaje, por lo que conlleva a que en paralelo también el facilitador va desarrollando, aprendiendo y evaluando bajo el enfoque de competencias, su propia praxis educativa.

Características del curso

El curso tal y como aparece en este manual, pretende abarcar los aprendizajes esenciales que le sean útiles al alumno del semestre correspondiente de bachillerato, en los horarios asignados por las autoridades directivas de cada plantel a los titulares de la asignatura. La modalidad del curso es a distancia, es decir, utilizando las herramientas digitales que le permitan al docente comunicarse en el marco de la presente contingencia por la pandemia e interactuar con sus alumnos no teniéndolos presentes físicamente. No obstante, considerando que existen alumnos que no cuentan con los recursos y/o conectividad para sus clases en línea, este manual va dirigido a ellos principalmente.

Los manuales están estratégicamente diseñados para propiciar un aprendizaje autodidacta para quienes no cuentan con los recursos y/o conectividad, así como la participación activa de quienes llevan sus clases en la modalidad en línea, la cual implica un compromiso entre el facilitador y los alumnos para alcanzar los objetivos del curso. Asimismo, las etapas de apertura, desarrollo y cierre, así como las actividades de contextualización y transversalidad y el tipo de ejercicios, permitirá crear las condiciones para estimular un trabajo en el que prevalezca la intención comprometida de cada uno de los participantes, para analizar y extraer las características más relevantes de las situaciones

problemáticas; discutir y encontrar formas de solución de los problemas y elegir, entre ellas, las más eficaces, así como fundamentar, en todo momento, el porqué de la estrategia de solución.

Un escenario de este tipo pretende crear las condiciones que propician aprendizajes significativos desde la distancia, donde lo más importante radica en ser consciente de lo que se hace y para qué se hace, y no sólo de solucionar el problema. En esta perspectiva, el docente está comprometido a supervisar de manera permanente el trabajo de sus alumnos, orientar y retroalimentar los contenidos que se requieran en plenarias, o en especial individualización, respetando los procesos de discusión y los argumentos que conduzcan al entendimiento y solución de los ejercicios, atender las dudas individuales y propiciar, siempre, la participación activa y comprometida de los estudiantes. Asimismo, el titular deberá realizar las siguientes actividades:

1. Al inicio del curso, el facilitador creará la herramienta digital que considere pertinente (Zoom, Google Meet, Classroom, WhatsApp, correo electrónico, etc.) y cerciorarse que esté incluida la totalidad de sus alumnos en sus grupos escolares correspondientes.

2. Ya creados los grupos digitales, realizar una dinámica para tratar de conocer a sus alumnos y explicar los objetivos del curso, duración, desarrollo, evaluación y compromisos que se adquieren al asistir al mismo.

3. Podrá hacer uso de la metodología del aula inversa a través de videos que ilustren el desarrollo de las actividades a realizar en cada sesión del curso. Dichos videos han sido seleccionados de la plataforma *Khan Academy* y *YouTube* y serán analizados por los alumnos el día anterior como una actividad extra clase a la sesión correspondiente de cada uno de los temas.

4. Apertura de sesiones. Se recomienda que la apertura se realice con el pase de lista y la resolución de la tarea diaria. Retroalimentando los errores identificados y aclarando dudas.

5. Cierre de sesiones. El cierre se realizará con una pregunta y los comentarios que de ella se deriven. Las preguntas pueden ser: ¿Qué aprendimos el día de hoy? ¿Cuál fue el error más grave que cometimos y cómo lo resolvimos?, entre otras.

6. Asesoría y seguimiento del desempeño de alumnos en la resolución de ejercicios para el aprendizaje y habilidad matemática, marcando un tiempo para su realización individual, al término del cual se preguntará quiénes han concluido, socializando en plenaria las soluciones.

7. Incluir en clase los Retos Transversales y las lecciones Construye T correspondientes a la asignatura y desarrollarlos, considerando una calificación ponderada formativa y sumativa.

8. Considerando la situación especial de contingencia por la pandemia del Covid 19, se podrá omitir la coevaluación y autoevaluación de los alumnos. Al término del curso, el docente evaluará en una escala de 0 a 10, los siguientes aspectos:

· Cumplimiento de los objetivos del curso.
· Dominio de los contenidos.
· Cumplimiento individual de las tareas, trabajos y evaluaciones en tiempo y forma. (Salvo casos especiales y no conectados)

9. Esta obra se hará llegar a los alumnos por los medios que dispongan en el contexto de cada región del país, tratando de abarcar la totalidad de la población de estudiantes de la DGETI. Para ello, en los planteles se establecerán los mecanismos para que se lleve a cabo una interacción favorable entre maestros y alumnos, a fin de dar seguimiento a los avances que tengan los jóvenes y establecer los criterios de evaluación que se consideren viables de acuerdo con las circunstancias de cada región, en el marco de la contingencia actual.

Recomendaciones para la impartición del curso

Este material contempla en su estructura una serie estrategias didácticas y ejercicios con un grado de complejidad gradual ascendente, cuyo principal propósito es que los procedimientos para su resolución y respuestas sirvan de parámetro a todos los involucrados en el proceso educativo, para emitir una opinión basada en el análisis de su alcance e importancia de desarrollarse siguiendo un razonamiento lógico-matemático.

Debido a la trascendencia académica del curso sugerimos tomar en cuenta las siguientes recomendaciones:

1. Los ejercicios tienen un grado de complejidad ascendente, por lo que es recomendable que el docente informe a los alumnos sobre el impacto que tiene cada habilidad en el aprovechamiento escolar; de igual forma es pertinente que si observa en el grupo dificultades en alguna habilidad la ejercite hasta que se domine, o en su defecto, brinde la oportunidad al estudiante de desarrollarla en otro espacio (plataforma Khan Academy, por ejemplo), o la estrategia que el considere pertinente.

2. El docente podrá grabar sus propios videos explicativos, proporcionar links de videos y textos explicativos de los temas, tutoriales, etc. con el propósito de que el estudiante tenga los recursos suficientes para la adquisición de las competencias y aclaración de posibles dudas en los contenidos.

5. Proporcionar al alumno y si es posible a los padres de familia (grupo de Whats App), los aspectos a considerar en la evaluación y su promedio parcial y final a tiempo para que tenga oportunidad de prepararse y regularizarse, de ser necesario.

6. Se debe tener consideración y empatía con aquellos alumnos que no tengan el recurso de conectarse diariamente y tratar de localizarlos con medios que estén al alcance de sus posibilidades y dándoles la oportunidad de trabajar o regularizarse en las condiciones que le favorezcan. Como, por ejemplo, ponerse de acuerdo en entregar tareas o evaluaciones en un punto de reunión física, por excepción y siguiendo las consideraciones de la contingencia.

Competencias a desarrollar en el curso

COMPETENCIA	ATRIBUTOS
1. Se conoce y valora así mismo y aborda problemas y retos teniendo en cuenta los objetivos que persigue.	1. Enfrentan las dificultades que se le presentan y es consciente de sus valores, fortalezas y debilidades.
	2. Identifica sus emociones, las maneja de manera constructiva y reconoce la necesidad de solicitar apoyo ante una situación que lo rebase.
4. Escucha, interpreta y emite mensajes pertinentes en distintos contextos, mediante la utilización de medios, códigos y herramientas apropiadas.	1. Expresa ideas y conceptos mediante representaciones lingüísticas, matemáticas o gráficas.
	2. Aplica distintas estrategias comunicativas según quienes sean sus interlocutores, el contexto en que se encuentra y los objetivos que persigue.
5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos.	1. Sigue instrucciones y procedimientos de manera reflexiva, comprendiendo cómo cada uno de sus pasos contribuye al alcance de un objetivo.
	6. Utiliza las TIC para procesar e interpretar información.
8. Participa y colabora de manera efectiva en equipos diversos.	2. Aporta puntos de vista con apertura y considera los de otras personas de manera reflexiva.
	3. Asume una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos grupos de trabajo.

Introducción

Debido a la presente prolongación del confinamiento social por la pandemia del SARS-CoV-2, una vez más las autoridades de la Secretaría de Educación Pública de México, han optado por la apertura de las clases a distancia en todos los niveles educativos, aprovechando los medios electrónicos actuales para que los docentes puedan desarrollar su cátedra de manera digital, teniendo comunicación con sus grupos de alumnos y así poder desarrollar las estrategias pertinentes que le permitan al estudiante alcanzar, en lo mayor posible, las competencias establecidas en los planes y programas de estudio nacionales.

Este manual representa la segunda edición (el primer manual fue generado de manera emergente por la Academia Nacional de Matemáticas, al iniciar la suspensión de clases presenciales en el país, en marzo del 2020 y se enfocó en su momento en cubrir los aprendizajes esenciales del segundo y tercer periodo parcial del semestre que quedaban pendientes, ya que el primer parcial se alcanzó a realizar de manera presencial en las aulas). Para ésta edición se incorporan las actividades propuestas para el primer periodo parcial, determinados por la Academia Nacional de Matemáticas, con el propósito de establecer los Aprendizajes Esenciales que se requieren de ésta asignatura en la formación de los alumnos de bachillerato de la DGETI.

Se trata de una estrategia didáctica que les permitirá a los estudiantes de bachillerato de este subsistema, tanto para los que cuentan con recursos para la modalidad en línea o para los que no, adquirir las competencias necesarias a partir de la recuperación de los conocimientos previos y la construcción de aprendizajes elementales, para continuar con su desarrollo y formación académica a través de la adquisición del sentido numérico, con el cual pueda transitar eficientemente hacia el manejo y comprensión de la abstracción que da el conocimiento lógico-matemático. La construcción del conocimiento deberá ser individual y colaborativa, donde todos los estudiantes tengan la oportunidad de adquirir los mismos conocimientos, según su propia percepción de la realidad.

El curso consta de tres periodos parciales, donde el alumno, guiado por el docente titular, deberá participar activa y dinámicamente en la construcción de sus aprendizajes y la solución de problemas en cada asignatura, en el marco de un ambiente digital y a distancia, debido a la imposibilidad de realizarse presencialmente por el riesgo de contagios presente en esta época de pandemia que nos tocó vivir.

El manual está estructurado en secciones que incluyen actividades de apertura, desarrollo y cierre como estrategias sistemáticas que le permitan al estudiante construir su conocimiento personal, adueñándose del manejo de las herramientas esenciales que le serán útiles en la adquisición de conocimientos formales posteriores y llegar a alcanzar su formación profesional y poder intervenir en los cambios que la sociedad actual le demande.

¡Somos orgullosamente DGETI!

Justificación

Si bien es cierto, las dificultades de comprensión y habilidades en matemáticas no se generan en el bachillerato, pero sí se reflejan en el aprovechamiento de los alumnos en este nivel y por consecuencia en la educación superior, por lo que se hace necesario emprender acciones dirigidas a subsanar dichas inconsistencias. Estamos convencidos que los jóvenes de nuevo ingreso al nivel medio superior mejorarán con la práctica su capacidad de observación, globalización, jerarquización, regulación de su propia comprensión, y por consecuencia, sus competencias matemáticas, cuya utilidad se verá reflejada, no sólo en el contexto académico, sino en cualquier ámbito de su vida cotidiana. Para los estudiantes que ingresan al bachillerato, es importante que inicien con una recapitulación de sus estudios básicos, porque el conocimiento de los números es una herramienta indispensable para comprender los procesos y fenómenos sociales y naturales, además es el fundamento para iniciar con los procesos de abstracción que requiere el álgebra, la geometría y el cálculo.

Bloque 1 | Manejo de la información

1.1 Definición de Estadística

1.1.1 Campo de estudio.

Actividades de Apertura

"Todas las oscuridades que presenta nuestra sociedad podrían ser iluminadas por la Estadística". Günter Grass. Premio Nobel 1999

La estadística juega un papel fundamental en las sociedades actuales en las que se producen y utilizan grandes cantidades de información, en este sentido, tiene que ver con las formas en que se recogen, organizan y comunican conjuntos de datos y con la manera en que éstos se analizan para hacer inferencias y predicciones para la toma de decisiones.

La estadística extiende su radio de influencia a todas las disciplinas científicas y sociales ya que como afirma *Moore (2000)* proporciona un método general útil para tratar datos, estimar su variación y el riesgo en situaciones de azar, de modo que puede decirse que, es tan importante en la cultura de las personas, que forma parte de un bagaje básico de alfabetización que todos debieran tener.

La estadística es un método científico que encuentra aplicación en una gran diversidad de campos del saber humano y cuya utilidad, como quedó demostrado desde el siglo pasado, va más allá de la mera descripción, pues permite el descubrimiento de leyes y tendencias. Dentro de los muchos ejemplos que permiten ilustrar esto, basta con citar el caso del estadístico alemán Ernesto Engel** (1821 – 1896) que adquirió renombre en el terreno de las investigaciones económicas y sociales al descubrir la ley que lleva su nombre y que se enuncia así: “Cuánto menor es el ingreso familiar, mayor es la proporción destinada a la compra de alimentos”.

Con datos recabados en 1857, observó que esa proporción era de 62%, 55% y 50% en familias de clase baja, media y alta, respectivamente. Al difundirse esta ley, resultó evidente que cuanto mayor es la parte del ingreso familiar que se invierte en alimentos, menor es la que se puede destinar a otros fines (vestido, salud, recreación, comodidades, etc.) y viceversa. Por esta razón, esta proporción ha sido utilizada como unidad de medida del bienestar social.

No obstante, como todo instrumento, la estadística tiene sus limitaciones; no puede, por ejemplo, diseñar investigaciones ni seleccionar problemas para someterlos a estudio, ni puede, por si sola, aportar resultados valiosos o dar interpretaciones de resultados en ausencia de una sólida teoría. Por otro lado, todos los resultados estadísticos, exactos o no, expresan de modo preciso mediante números. Pero preciso no es sinónimo de exacto; son exactas las operaciones aritméticas, pero las mediciones que conducen a los datos que las hacen posibles no siempre son confiables. Por esta razón los resultados estadísticos deben ser siempre sometidos a crítica.

Pero ¿qué es la estadística? Desde mediados del siglo XVIII hasta una centuria después, la estadística ha sido objeto de muchísimas definiciones”, sin embargo, será suficiente por ahora con que nos familiaricemos con una definición que responde a los objetivos de este curso:

Actividades de Desarrollo

Definición de “Estadística”.

Ve al siguiente sitio https://www.ecured.cu/Estad%C3%ADstica y a partir del mismo responde en una cuartilla a las siguientes preguntas:

¿Qué es la estadística?

¿Qué campos conforman la estadística?

¿Qué métodos utiliza y en qué consisten?

¿Con qué disciplinas se relaciona?

Una vez que hayas terminado de dar respuesta a las preguntas, comparte comentarios al respecto de las preguntas con al menos un compañero y hagan llegar su trabajo a su profesor(a).

Habrás observado que se puede entender la estadística como la ciencia que permite responder a ciertas preguntas basándose en datos empíricos, es decir, en datos que se originan de la observación o la experiencia Entendida así, diremos que es la ciencia que tiene que ver con los métodos que dan respuesta a determinadas cuestiones, mediante la recolección y la interpretación apropiadas de datos empíricos. Las observaciones o las experiencias que constituyen los datos pueden resultar de la investigación científica, de la actividad comercial o de la vida cotidiana. En cualquier caso, la estadística busca dar sentido a los datos; esto implica tanto la recolección como la interpretación de éstos.

La recolección abarca el diseño de las investigaciones empíricas, la planeación de lo que se quiere observar, la calidad y suficiencia de la observación y el registro de los datos; la interpretación, el análisis y el resumen de los datos, la extracción de conclusiones a partir de ellos y el reporte y la presentación de los resultados.

De igual modo, habrá observado que, para su estudio, la estadística se divide en dos grandes ramas; descriptiva e inferencial.

Estadística Descriptiva
Es un conjunto de procedimientos que sirven para organizar, describir y sintetizar datos, sin que las conclusiones que se extraigan de éstos rebasen su ámbito específico.

Estadística Inferencial
Es un conjunto de procedimientos que se emplean para hacer inferencias y generalizaciones respecto a una totalidad, partiendo del estudio de un número limitado de casos tomados de esta última.

Las inferencias y generalizaciones en esta rama, que complementa a la descriptiva, se basan en la teoría de la probabilidad.

El carácter propio del método estadístico descansa en el estudio de grupos o masas, a través de los elementos que los componen. En estadística no interesan aisladamente las características de un

elemento de la masa. No interesa, por ejemplo, que la vida útil de una lámpara de cierto diseño sea de 10 mil horas y la de otra de 3 mil. Lo que importa es ver la tendencia de cierto número de lámparas que puedan ser representativas de toda la producción; lo que se busca es descubrir, por ejemplo, que la vida útil promedio de ese tipo de lámparas de 7 mil horas. Si lo que se estudia es un grupo de personas, no interesa que una de ellas en particular profese el catolicismo y otra el protestantismo, por mencionar algo; lo que quisiéramos conocer podrían ser los cultos existentes y el que más se profesa en el grupo.

El medio empleado para el estudio estadístico es la enumeración o recuento. Enumerar es captar las características de los elementos sometidos a estudio y anotarlos o medirlos bajo las condiciones que se presentan. La estadística es, básicamente, un método de inducción basado en los grandes números y sus propiedades, con lo cual se eliminan los errores propios de la observación y se aumenta la validez de los resultados obtenidos.

1.1.2 Antecedentes históricos.

La palabra estadística se origina, en las técnicas de recolección, organización, conservación, y tratamiento de los datos propios de un estado, con que los antiguos gobernantes controlaban sus súbditos y dominios económicos. Estas técnicas evolucionaron a la par con el desarrollo de las matemáticas, utilizando sus herramientas en el proceso del análisis e interpretación de la información.

A modo de una brevísima nota histórica se menciona que; para mediados del siglo XVII en Europa, los juegos de azar eran frecuentes, aunque sin mayores restricciones legales. El febril jugador De Meré consultó al famoso matemático y filósofo Blaise Pascal (1623-1662) para que le revelara las leyes que controlan el juego de los dados, el cual, interesado en el tema, sostuvo una correspondencia epistolar con el tímido Pierre de Fermat (1601-1665, funcionario público apasionado por las matemáticas; célebre porque no publicaba sus hallazgos) dando origen a la teoría de la probabilidad, la cual se ha venido desarrollando y constituyéndose en la base primordial de la estadística inferencial.

Hoy en día, son de uso cotidiano las diferentes técnicas estadísticas que partiendo de observaciones muéstrales o históricas, crean modelos lógico-matemáticos que se "aventuran" describir o pronosticar un determinado fenómeno con cierto grado de certidumbre medible.

El presente material no pretende teorizar el saber estadístico, desde luego, no es un material para estadísticos, ya que, adrede se obvia el rigor científico de lo expuesto en beneficio de la sencillez necesaria para el neófito; con un lenguaje coloquial se conduce al estudiante a través del contenido, a partir de dos o tres ejemplos que ilustran la aplicabilidad de los temas tratados.

La estadística, entonces, dejó de ser una técnica exclusiva de los estados, para convertirse en una herramienta imprescindible de todas las ciencias, de donde proviene la desconcertante des-uniformidad en las definiciones de los diferentes autores, ya que cada estudioso la define de acuerdo con lo que utiliza de ella, se tienen definiciones como que: la estadística es la tecnología del método científico, o que es el conocimiento relacionado con la toma de decisiones en condiciones de incertidumbre, o que la estadística son métodos para obtener conclusiones a partir de los resultados de los experimentos o procesos, o que es un método para describir o medir las propiedades de una población. En fin, no se trata de discutir si la estadística es una ciencia, una técnica o una herramienta, sino de la utilización de sus métodos en provecho de la evolución del conocimiento.

La estadística hace inferencias sobre una población, partiendo de una muestra representativa de ella. Es a partir del proceso del diseño y toma de la muestra desde donde comienzan a definirse las

bondades y confiabilidad de nuestras aseveraciones, hechas, preferentemente, con un mínimo costo y mínimo error posible.

En las últimas décadas la estadística ha alcanzado un alto grado de desarrollo, hasta el punto de incursionar en la totalidad de las ciencias; inclusive, en la lingüística se aplican técnicas estadísticas para esclarecer la paternidad de un escrito o los caracteres más relevantes de un idioma.

La estadística es una ciencia auxiliar para todas las ramas del saber; su utilidad se entiende mejor si se tiene en cuenta que los quehaceres y decisiones diarias embargan cierto grado de incertidumbre... y la Estadística ayuda en la incertidumbre, trabaja con ella y nos orienta para tomar las decisiones con un determinado grado de confianza.

Los críticos de la estadística afirman que a través de ella es posible probar cualquier cosa, lo cual es un concepto profano que se deriva de la ignorancia en este campo y de lo polifacético de los métodos estadísticos. Sin embargo, muchos "investigadores" tendenciosos han cometido abusos con la estadística, elaborando "investigaciones" de intención, teniendo previamente los resultados que les interesan mostrar a personas ingenuas y desconocedoras de los hechos. Otros, por ignorancia o negligencia, abusan de la estadística utilizando modelos inapropiados o razonamientos ilógicos y erróneos que conducen al rotundo fracaso de sus investigaciones.

Lincoln L. Chao (Estadística para ciencias administrativas. 3a ed. México: Mc Graw Hill 1993) hace referencia a uno de los más estruendosos fracasos, debido a los abusos en la toma de una muestra:

Se trata del error cometido por la Literary Digest que, en sus pronósticos para las elecciones presidenciales en EE. UU. para 1936, afirmó que Franklin D. Roosvelt obtendría 161 votos electorales y Alfred Landon, 370. La realidad mostró a Roosvelt con 523 votos y a Landon con 8 solamente.

El error se debió a que la muestra fue tomada telefónicamente a partir de la lista de suscriptores de la Digest y, en 1936, las personas que se daban el lujo de tener teléfonos y suscripciones a revistas no configuraban una muestra representativa de los votantes de EE.UU. y, por ende, no podía hacerse un pronóstico confiable con tan sesgada información.

Ante todo, la estadística es un elemento útil, ya que permite obtener información referida a grandes grupos de individuos conociendo los datos de sólo unos pocos.

El desarrollo de la Estadística se fundamenta científicamente a partir de los años 30 a raíz de los problemas planteados en la sociedad industrial, por el desarrollo de otras ramas de las Matemáticas y de otros campos como la Biología, Medicina, Informática...

Actualmente, la estadística se ha convertido en un método efectivo para describir con exactitud los valores de datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y sirve como herramienta para relacionar y analizar dichos datos. El trabajo del experto estadístico no consiste ya sólo en reunir y tabular los datos, sino sobre todo en el proceso de interpretación de esa información. El desarrollo de la teoría de la probabilidad ha aumentado el alcance de las aplicaciones de la estadística. Muchos conjuntos de datos se pueden aproximar, con gran exactitud, utilizando determinadas distribuciones probabilísticas; los resultados de éstas se pueden utilizar para analizar datos estadísticos. La probabilidad es útil para comprobar la fiabilidad de las inferencias estadísticas y para predecir el tipo y la cantidad de datos necesarios en un determinado estudio estadístico.

La sociedad tiene la necesidad de conocer toda esta información, y de acceder a ella de la forma más resumida posible.

Actividades de cierre

A partir de consultar el sitio https://www.ine.es/explica/docs/historia_estadistica.pdf responda a las siguientes preguntas;

¿Cómo surge la estadística?

¿Cuáles son los principales registros que se han documentado al respecto de la historia de la estadística, en particular antes de la era cristiana con las civilizaciones de los egipcios, romanos, chinos y griegos?

¿Qué dice la historia, del desarrollo de la estadística durante la edad media?

¿Qué se sabe del desarrollo de la estadística en la edad Moderna?

¿Qué se sabe del desarrollo de la estadística en la edad contemporánea?

¿Qué relevancia tendrá la estadística como para fijar el día mundial de la estadística?

Una vez que haya terminado de dar respuesta a las preguntas, haga llegar su trabajo a su profesor(a)

1.2 Población, muestra y tipos de variables.

Introducción

La mayoría de las ciencias avanza y crea marcos teóricos, comprobando con la observación directa de todo lo conocido. Un ejemplo es el de las aves, hasta hace unos cientos de años los biólogos tenían a un conjunto de animales que tenían pico, todos ellos eran aves, ya sabes; con plumas, dos patas, forma conocida, etc. En base a ellas se crearon libros, categorías, cursos completos. Cuando se conoció para el mundo europeo–americano al ornitorrinco, la mayoría de las teorías al respecto se vinieron abajo.

Con lo anterior queremos establecer que, para hacer una afirmación científica absoluta, deberíamos conocer a TODOS los elementos que vamos a estudiar, esto es en ocasiones imposible, de ahí que se han desarrollado técnicas para el análisis de una parte de los elementos a estudiar.

Actividades de Apertura

Reflexiona lo siguiente: Si quisiéramos saber el número de habitantes de un estado o municipio, tardaríamos varias semanas en contar a todos y cada uno de ellos, cuando finalmente terminamos, ya nacieron o fallecieron algunos de sus habitantes.

Es por ello que se hacen estudios de una porción de los habitantes y luego se generaliza. Esto también es aplicable a la industria en general.

Actividades de Desarrollo

1.2.1 Población

La población o Universo es todo conjunto de personas, cosas, objetos, etc., con ciertas características comunes que puede ser finita o infinita.

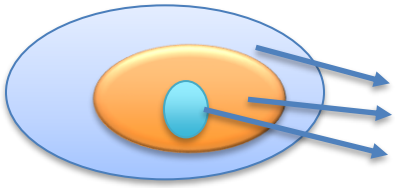
1.2.2 Muestra

La **Muestra** es toda porción de elementos tomada de una población, entre más grande sea la muestra, será más representativa de la población.

El **Muestreo** es el procedimiento por el cual se recopila información de los elementos de una muestra, existen varias formas de realizar el muestreo, unas son más sencillas que otras, otras más económicas y otras más representativas.

Ejemplo 1:

Se desea estudiar las características económicas de un estado de la República Mexicana. Escribe tu propuesta de un ejemplo de Universo, Muestra y elemento.



Universo	País – Habitantes de México
Muestra	Estado – Habitantes de Michoacán
Elemento	Municipio – Habitantes de Jiquilpan

Ejemplo 2:

Se desea estudiar las Socio económicas de los alumnos del CBTis 12 de Jiquilpan, Mich. Para facilitar el estudio, se decide tomar como muestra a un grupo al azar. Escribe tu propuesta de un ejemplo de Universo, Muestra y elemento.

Universo:	Alumnos del CBTis 12
Muestra:	Alumnos de 5° “A” de Laboratorista.
Elemento:	Joaquín Díaz Juárez de San Diego Quitupán.

1. Escribe 2 ejemplos que contengan universo, población y muestra. Pueden ser de artículos de tu agrado como: celulares, motos, consolas de juegos, etc.

	Ejemplo 1	Ejemplo 2
Universo		
Muestra		
Elemento		

1.2.3 Tipos de variables

Una **Variable** se puede representar con un símbolo tal como x, y, a, b, que puede tomar un valor cualquiera. Si esta variable solamente toma un valor, se le llama **Constante**.

Alumno	Promedio
Francisco	9.2
Felipe	8.7
Raúl	9.5
Jaime	8.0

Variable

En este caso el promedio es la variable.

1.2.3.1. Variables cualitativas

Las variables cualitativas son aquellas que describen las cualidades o características de un objeto o evento mediante palabras, frases, letras, etc. Son subjetivas de acuerdo con el observador. Por

ejemplo, si deseamos describir el estado de conservación de una bicicleta, alguien podría decir que su estado es muy bueno, mientras que otra la calificaría como Bueno.

Las variables cualitativas se clasifican en:

Variables Nominales son las más simples y abundantes, su única función es la de clasificar en categorías y su orden es indistinto. NO tiene sentido realizar operaciones aritméticas con ellas.

Ejemplo

	Estado Civil				
Variable →	Soltero	Casado	Divorciado	Viudo	Unión libre
Categoría →	1	2	3	4	5

Aquí los números carecen de propiedades solo sirven para distinguir un estado civil de otro, no significa que un estado civil sea mejor o mayor que otro.

Variables Ordinales clasifican las observaciones en categorías que exigen ordenación, su variable operacional es una escala ordinal de mayor a menor.

Ejemplo

	Grado de Alcoholismo			
Variable →	Abstemio	Ocasional	Regular	Constante
Escala de medición →	1	2	3	4

NO tiene sentido realizar operaciones aritméticas con ellas.

1.2.3.2. Variables cuantitativas

Las variables cuantitativas son aquellas que se pueden representar por una cantidad numérica exacta, por lo general son medidas objetivas, es decir, que no depende de quien realice la medición. En ocasiones puede tener un pequeño error dependiendo del instrumento con que se mida.

Las variables cuantitativas se clasifican en:

1.2.3.3. Variables Discretas

Variables Discretas son las que toman solo algunos valores dentro de un intervalo, por ejemplo, valores enteros. Al indicar el número de hijos que tiene una familia, el número de recámaras de una casa, el número de mascotas, etc.

Ejemplo

Variable →	Número de hijos por familia						
Escala de medición →	1	2	3	4	5	6	7 o más

1.2.3.4. Variables Continuas

Variables Continuas Son las que pueden tomar cualquier valor fijo dentro de un intervalo, siempre entre dos valores observables, va a existir un tercer valor intermedio, que también podría tomar la variable continua.

Ejemplo

Variable →	Estatura (escuela primaria)						
Escala de medición →	1.10m	1.2m	1.21m	1.215m	1.35m	1.352m	1.40m

Una variable continua toma valores a lo largo de un continuo, esto es, en todo un intervalo de valores. Un atributo esencial de una variable continua es que, a diferencia de una variable discreta, nunca puede ser medida con exactitud; el valor observado depende en gran medida de la precisión de los instrumentos de medición. Con una variable continua hay inevitablemente un error de medida.

Actividades de cierre

Reflexiona cuando en alguna ocasión te ha tocado llenar un formulario, ya sea de inscripción, de apertura de una cuenta, una solicitud, etc. ¿Cuál tipo de variable te recordó al ver las anteriores?

Llena un ejemplo de ese tipo de variable que recuerdas y que sea distinto a los ejemplos ya vistos.

Ejemplo de una variable _____

Variable →							
Escala de medición →							

1.3. Niveles de medición de variables.

Actividades de Apertura

Se define como variable a la característica, cualidad o atributo de interés de las unidades de una población o una muestra.

Si las unidades de interés son estudiantes, las variables que pueden resultar relevantes son: edad, promedio general de aprovechamiento, estatura, peso, sexo, entre muchos otros; si las unidades de estudio son frutas, las variables son: pesos, sabor, color, textura, contenido de agua, valor proteínico o vitamínico, etcétera.

Las variables pueden tomar diferentes valores, los cuales se clasifican en las escalas siguientes:

Actividades de Desarrollo

1.3.1. Nominal

Esta escala utiliza para clasificar a la población en diferentes categorías, en las cuales no existe una forma particular para ordenar las etiquetas. Por ejemplo, un ser humano puede ser de sexo masculino o femenino; la clasificación de los seis colores de las lunetas de chocolate de leche M&M; los ciudadanos pueden tener como estado civil: soltero o casado.

1.3.2. Ordinal

La escala ordinal se caracteriza porque las modalidades que integran a la misma tienen un orden entre ellas. Se distingue porque no se define la distancia entre las observaciones. Por ejemplo, México es un país que consume altas cantidades de refresco, pero al entrevistar a una persona acerca de la frecuencia con que consume esta bebida, se pueden tener respuestas como: regularmente, algunas veces, pocas veces, casi nunca, nunca; si se cuestiona a los ciudadanos de un municipio si están de acuerdo con la forma de gobierno, se podrán tener respuestas como: totalmente de acuerdo, de acuerdo, indiferente, en desacuerdo y totalmente en desacuerdo.

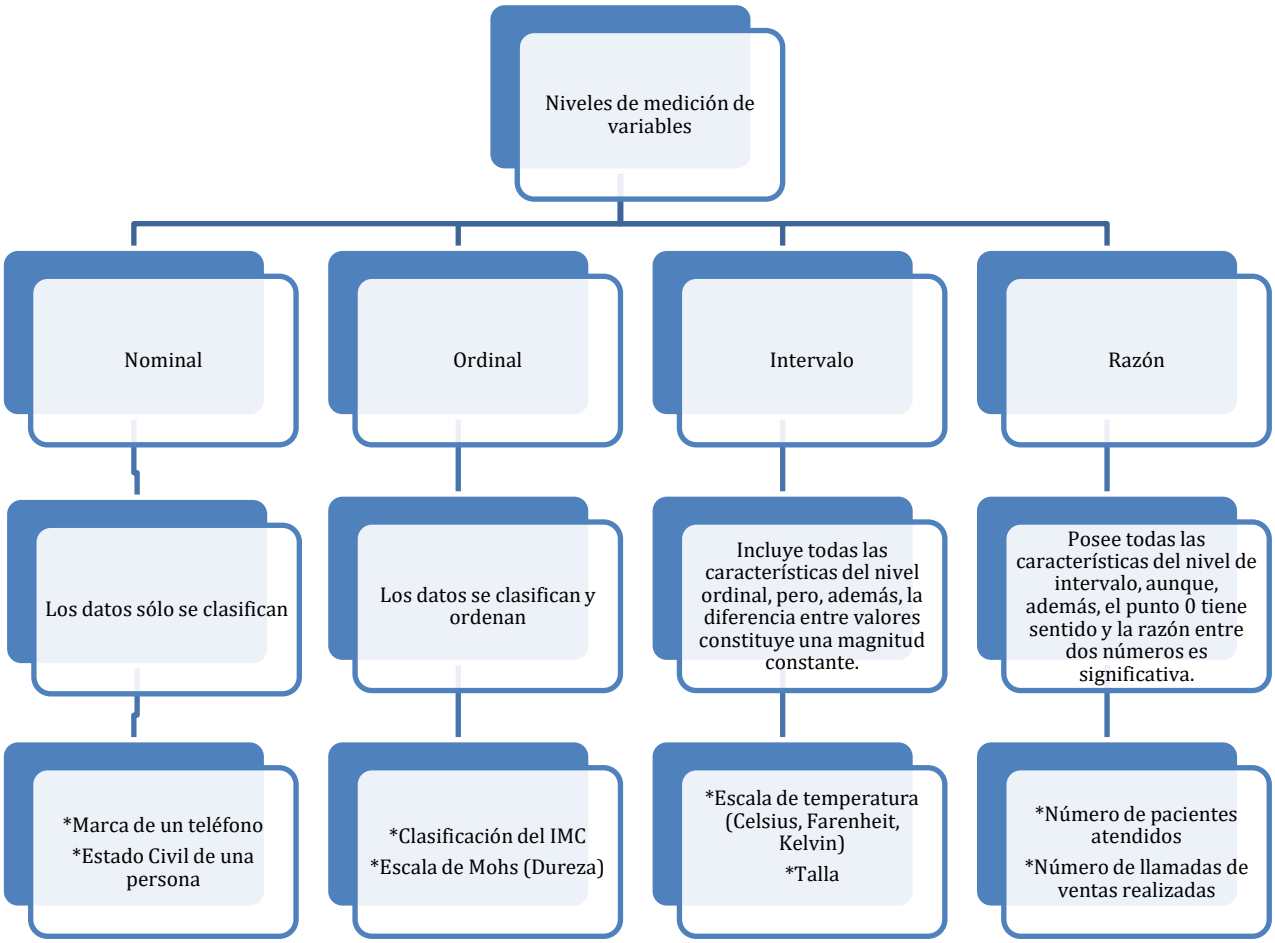
1.3.3. De intervalo

Esta escala, además de clasificar y ordenar, establece la proporción o la distancia entre dos intervalos contiguos. Requiere una unidad de medida y un punto cero arbitrario (no necesariamente el cero absoluto). Un ejemplo clásico de este tipo de escala es la temperatura en grados centígrados, Kelvin o Fahrenheit, ya que la unidad de medida y el punto cero son arbitrarios.

1.3.4. De razón

Este tipo de escala clasifica, ordena, establece la proporción o distancia entre dos intervalos contiguos y agrega el cero absoluto; además son permisibles las operaciones aritméticas (suma, resta, multiplicación y división) e implica las relaciones de equivalencia (mayor que, menor que, y equivalente). Hay muchos casos aplicables a esta escala; por ejemplo, se puede cuestionar a un grupo de personas sobre sus pesos o pesar diferentes objetos; es posible estimar la altura de diferentes objetos, animales o personas.

La siguiente gráfica resume las principales características de los diversos niveles de medición.

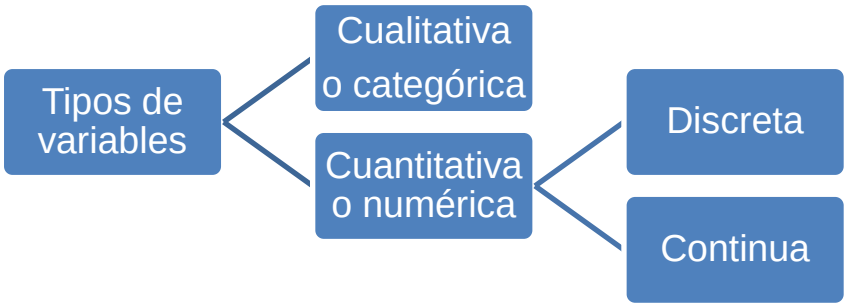


Cuando se mide una variable, el resultado de la medición se puede expresar con base en sus características o de forma numérica. Por tanto, se tienen dos tipos de variables:

Cualitativa o categórica. Cuyos posibles valores son únicamente categorías. En este tipo de variables se ubican las categorías nominal y ordinal.

Cuantitativa o numérica. Cuyos posibles valores son números, en ésta se ubican las categorías de intervalo y razón.

La siguiente gráfica resume los dos tipos de variables:



1. Señala con una palomita el tipo de escalas de medida a la que pertenece cada una de las variables siguientes:

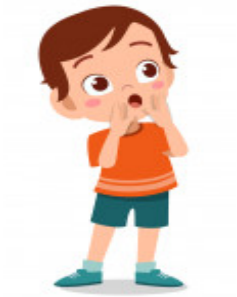
No.	Variables	Nominal	Ordinal	De intervalo	De razón
1	Peso de una persona				
2	La raza de un perro				
3	Frecuencia cardiaca				
4	Grado de una enfermedad				
5	Marca de un automóvil				
6	La escolaridad de una persona				
7	Nacionalidad				
8	Cantidad de Hemoglobina				
9	Lugar de nacimiento				
10	Número de hijos				

Actividades de cierre

1. En tu cuaderno, elabora un mapa conceptual de los tipos de variables y las categorías de los datos. En la elaboración de tu mapa incluye algunos ejemplos.

1.4 Frecuencias

Introducción



Agrupar datos y clasificarlos son tareas importantes para producir información. En las tablas estadísticas de frecuencias, cada clase tiene una frecuencia de la clase, que es el número de mediciones que se pueden agrupar en ella. Con las clases y sus frecuencias se construye la tabla de frecuencias, también llamada distribución de frecuencias, de la que se obtiene la frecuencia relativa de cada clase.

Actividades de Apertura

La frecuencia estadística es la cantidad de veces que se repite una observación durante la realización de un muestreo.

1.4.1. Frecuencia Absoluta (acumulada)

Se le llama frecuencia absoluta al **número de veces que se repite una variable en un experimento**. Esta se representa con f_i **ó** n_i.

Frecuencia absoluta acumulada

La frecuencia acumulada es aquella que se obtiene al sumar todas las frecuencias absolutas inferiores o iguales al valor en cuestión.

1.4.2. Frecuencia Relativa (acumulada)

Representa la cantidad de veces que se repite una observación, expresada como proporción de la muestra, es decir, es el resultado de dividir el valor de la frecuencia absoluta por el tamaño de la muestra estadística.

Esta se representa con f_r y se define como $f_r = n/N$, siendo **n** el número de veces que se repite la respuesta y **N** el tamaño de la muestra. Su valor se expresa como porcentaje.

Frecuencia relativa acumulada

En esta se tiene en cuenta la sumatoria de todas las frecuencias relativas inferiores o iguales al valor en cuestión. Se representa con f_r.

La frecuencia relativa acumulada de la clase i es igual a la suma de las frecuencias relativas de todas las clases anteriores a ella más la de la clase i.^{i}

Actividades de Desarrollo

1. Analiza los siguientes ejemplos desarrollados, ya que esto te permitirá poder utilizar esta información más adelante en tablas más complejas para encontrar las medidas de tendencia central o de dispersión.

Ejemplo 1.

En una autopista se midieron las velocidades de 50 automóviles en km/h utilizando una lámpara para medir la velocidad, y cuyos resultados se anotan a continuación.

88	106	107	145	142	91	103	87	130	99
132	144	112	89	137	110	139	105	147	119
123	108	133	111	122	123	94	145	113	146
109	108	88	120	138	91	135	145	110	95
100	137	110	137	142	120	113	140	136	150

Más adelante observarás la explicación para ordenar los datos, en este espacio te explicaré como retomar las frecuencias y colocarlas en la siguiente tabla.

Al ordenar los datos en esta tabla puedo empezar a retomar los datos que se repiten a esto le llamaremos frecuencias

Tabla 1. Velocidad de automóviles en una autopista.

Clase	Intervalo VELOCIDAD (Km/h)	Frecuencia absoluta $f_i = f$	Frecuencia Relativa f_r	Frecuencia Acumulada cf
1	87 - 95	8	0.16	8
2	96 - 104	3	0.06	11
3	105 - 113	13	0.26	24
4	114 - 122	4	0.08	28
5	123 - 131	3	0.06	31
6	132 - 140	10	0.2	41
7	141 - 149	8	0.16	49
8	150 - 158	1	0.02	50
		N=n=$\sum f$ =50	**1**	

Observa que la suma de la frecuencia relativa siempre será la unidad, en el caso de la frecuencia acumulada no se suma ya que esta acumulada dando el total de las frecuencias.

Ejemplo 2:

En una investigación acerca de la temperatura máxima en grados centígrados en la ciudad de Veracruz en la temporada de verano, durante 4 años se obtuvieron datos al azar y se tabularon como se muestra a continuación.

En la siguiente tabla pondrás en práctica los conceptos analizados.

Tabla 2. Temperaturas máximas durante el verano en la ciudad de Veracruz (muestras aleatorias de 4 años).

Clase	**Intervalo de clase: TEMPERATURA $°C$**	**Frecuencia absoluta** $f_i = f$	**Frecuencia Relativa** f_r	**Frecuencia Acumulada** cf
1	28 - 30	3	0.025	3
2	31 - 33	10	0.083	13
3	34 - 36	20	0.167	33
4	37 - 39	28	0.233	61
5	40 - 42	35	0.292	96
6	43 - 45	24	0.2	120
	Total	**n=120**	**1**	

Podrás observar que la frecuencia absoluta son las frecuencias obtenidas directamente de la información, en cuanto a la frecuencia relativa. La frecuencia absoluta es una medida estadística que nos da información acerca de la cantidad de veces que se repite un suceso al realizar un número determinado de experimentos aleatorios, esta medida se representa cono ***fi o f,*** como ya se mencionó en párrafos anteriores, se refiere a la realización i-énesima del experimento aleatorio.

La frecuencia relativa es el resultado de dividir la frecuencia absoluta de un determinado valor entre el número total de datos registrados, la frecuencia absoluta relativa será el número de resultados de dividir, la frecuencia acumulada entre el número total de registros.

Actividades de cierre

En este caso se está agregando una columna para que agregues la frecuencia relativa acumulada.

Vida de una pieza metálica.

1. Una máquina para triturar piedra utiliza una pieza metálica sujeta a rotación y fricción. Esa pieza tiene lo que se llama un “tiempo” de vida útil”. El productor de las máquinas realiza frecuentemente pruebas a la pieza para conocer cuántas horas de duración de la pieza debe garantizarles a los clientes. En la siguiente tabla se muestran los resultados de la distribución de frecuencias de los tiempos de vida obtenidos en las 60 piezas.

Tabla 3. Duración de las piezas en una hora.

Clase	Intervalo de clase: HORAS (H)	Frecuencia f	Frecuencia Relativa f_r	Frecuencia Acumulada cf	Frecuencia Relativa Acumulada
1	2900 - 3100	5			
2	3101 - 3301	8			
3	3302 - 3502	12			
4	3503 - 3703	25			
5	3704 - 3904	10			
	Total				

Actividades de contexto o Transversales

1. Analiza los siguientes ejercicios como reforzamiento del tema analizado anteriormente Frecuencias absoluta y relativa, frecuencia acumulada y completa las tablas.

a) Los siguientes datos de la tabla muestra la distribución de frecuencias de tiempos de una muestra, tomada al azar de la duración en segundos de comerciales transmitidos por la televisión comercial en el horario de las 20:00 hrs. a las 23:00 hrs. tomada en un periodo de dos años. No se repite un comercial en la muestra.

Clase	Intervalo Real de clase: T	Frecuencia f	Frecuencia Relativa f_r	Frecuencia Acumulada cf	Frecuencia Relativa Acumulada
1	5< T ≤10	23			
2	10 <T ≤15	66			
3	15< T ≤20	49			
4	20< T ≤25	16			
5	25< T ≤30	4			
6	30< T ≤35	2			
	Totales	160			

b) Se realiza una investigación acerca del punto de fusión del plomo (Temperatura a la cual cambia de sólido a líquido). Se trabaja en un laboratorio con temperatura controlada a 26^oC en crisoles de acero (Fuente de fuego de precisión e instrumentos de medición de calidad). Los datos de una muestra de 100 observaciones se han vaciado en la tabla de distribución de frecuencias siguiente:

Clase	**Intervalo: Temperatura de fusión en ^{o}C**	**Frecuencia f**	**Frecuencia Relativa f_r**	**Frecuencia Acumulada cf**	**Frecuencia Relativa Acumulada**
1	326< C ≤326.4	8			
2	326.4< C ≤326.8	12			
3	326.8< C ≤327.2	70			
4	327.2< C ≤327.6	7			
5	327.6< C ≤328	3			
	Totales	n=			

1.5 Tablas de Distribución de Frecuencias

1.5.1 Elementos

Introducción

Los ***datos*** son situaciones o hechos que se representan numéricamente y que algunas veces forman parte de la vida cotidiana y otras, se encuentran en libros porque han sido recopilados por otras personas. Por lo anterior, los datos pueden ser:

Datos Originales: Son aquellos que son recopilados por el propio investigador; por lo tanto, son comprobables en forma rigurosa.

Datos Indirectos: Son aquellos que son recopilados de enciclopedias, libros de registro, sucesos grabados en audio y video, etc.

Una vez hecha la recopilación de los datos, el siguiente paso es su correcta organización para que brinden una información fiel y de gran utilidad, considerando su naturaleza de acuerdo con los propósitos para los que fueron recopilados.

Ordenación estadística de los datos: es el método o técnica que se usa para reunir datos numéricos, los cuales se pueden ordenar en forma creciente (ascendente) o decreciente (descendente).

Actividades de Apertura

1.5.1.1 Rango

El *Recorrido o Rango* es la diferencia que existe entre el mayor y menor de los datos:

$$\boldsymbol{R = X_M - X_m}$$

Ejemplos:

1. Ordena y calcula el rango de los siguientes datos: 8, 13, 29, 16, 32, 24, 17, 10, 21, 28, 35.

Solución:

Ordenación

(Ascendente) →
8, 10, 13, 16, 17, 21, 24, 28, 29, 32, 35
← (Descendente)

El rango se determina por la ecuación:

$$R = X_M - X_m$$
$$R = 35 - 8$$
$$R = 27$$

2. Las calificaciones finales de matemáticas de 40 estudiantes del CBTis No. 7 son:

95	70	96	85	100	49	83	89
55	55	65	77	80	70	92	93
74	66	95	65	87	100	45	77
60	75	69	52	82	68	78	92
58	56	70	70	74	98	75	64

Determina su ordenación Creciente y su Rango.

Solución:

La ordenación creciente es:

Dato menor	45	56	65	70	75	80	89	95	
	49	58	66	70	75	82	92	96	
	52	60	68	70	77	83	92	98	
	55	61	69	74	77	85	93	100	
	55	65	70	74	78	87	95	100	Dato mayor

Su rango es:

$$R = X_M - X_m$$
$$R = 100 - 45$$
$$R = 55$$

Ejercicios:

Desarrolla en equipo los siguientes ejercicios y discute en plenaria los resultados:

1. Ordena y calcula el Rango para los siguientes datos:

a) 68, 70, 74, 94, 75, 78, 88, 72, 63, 78, 85, 95, 77, 62, 93, 90.

b) 19, 54, 36, 29, 10, 48, 16, 52, 39, 24.

c) 2, 7, 20, 13, 1, 9, 16, 11, 5, 18.

2. Un cuestionario de 60 preguntas con el tema “Clásicos de la música”, aplicado a 50 ciudadanos comunes, dio lugar a los siguientes aciertos.

39	44	28	53	37	46	51	33	29	48
46	57	55	38	31	49	52	56	29	55
46	58	55	43	36	34	41	40	44	38
36	51	50	30	35	39	49	53	19	24
37	46	31	26	48	44	39	52	40	35

Considera los datos anteriores y determina:

a) Su ordenación Decreciente.
b) Su Rango.
c) Las cantidades de las 3 personas con el menor número de aciertos.
d) Las cantidades de las 3 personas con el mayor número de aciertos.
e) Las cantidades del primero al décimo lugar con mayor puntuación.
f) Número de personas que obtuvieron más de 40 aciertos.
g) Número de personas que obtuvieron menos de 30 aciertos.
h) La cantidad del vigésimo lugar en orden ascendente.

Actividades de Desarrollo

Distribución de Frecuencias de Datos

Cuando se tiene una gran cantidad de datos, se recomienda distribuirlos en clases o categorías y determinar con precisión el número de datos pertenecientes a cada clase. Este procedimiento se denomina *frecuencia de clase* o también, *frecuencia absoluta de clase*.

Lo anterior es de gran utilidad ya que permite analizar con mayor facilidad un grupo de datos sin que se tenga que considerar individualmente cada uno de ellos.

La orientación tabular de los datos en clases o categorías en la cual se conjunta cada clase con su respectiva frecuencia se denomina *distribución de frecuencias* o *tabla de frecuencias*.

Ejemplo:

La tabla siguiente es una distribución de frecuencias de los salarios diarios de 72 profesionales de la industria petrolera:

Tabla de frecuencias

Salarios (dólares)	Número de profesionales
30 - 39	7
40 - 49	12
50 - 59	19
60 - 69	16
70 - 79	10
80 - 89	6
90 - 99	2
	N= 72

Clases o Categorías

Frecuencias de clase

La segunda clase o categoría comprende los salarios de 40 a 49 dólares y se representa (40 – 49), dado que doce profesionales tienen un salario perteneciente a esta clase, la respectiva frecuencia absoluta es 12.

Cuando los datos se ordenan y se resumen en una tabla o distribución de frecuencias, éstos se denominan *datos agrupados*.

1.5.1.2 Clases o Intervalos

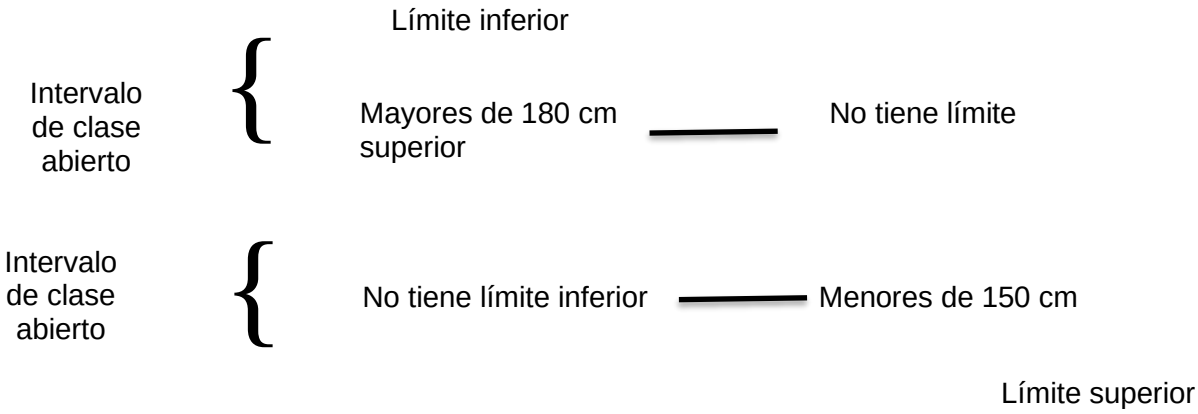
Si se observa la tabla del ejemplo anterior, la cuarta clase o categoría se denota por los números “60 – 69”. El conjunto de valores que abarcan desde 60 hasta 69 recibe el nombre de *intervalos de clase*.

Intervalo de clase abierto

Se define como el intervalo de clase abierto a aquel que, al menos teóricamente, no tiene límite inferior o límite superior.

Ejemplo:

Al referirse a las estaturas de grupos de individuos, los intervalos de clase que contienen estaturas mayores de 180 cm o menores de 150 cm, respectivamente, se consideran intervalos de clase abierto, es decir:



Los *intervalos de clase* o *intervalo exacto* son aquellos donde pueden agruparse los datos de una variable estadística. Es conveniente que cuando se seleccione el número de intervalos o de clases NC, éste no sea muy pequeño o muy grande, esto depende de la experiencia del analista que realice la tabla. En ocasiones, se toma el criterio de la raíz cuadrada del número de datos (N), en otras la raíz cúbica o la regla de Storges qué está dada por la ecuación:

$$\boldsymbol{NC = 1 + 3.3n}$$

También se puede seleccionar el número de intervalos o de clases dividiendo el rango entre un número cualquiera de clase, empezando desde 5 y hasta 20; se selecciona el número de clase en el cual el resultado de la división (anchura) sea un número que se puede redondear al inmediato superior.

1.5.1.3 Límites o fronteras de clase

Cada intervalo de clase delimita los valores que ésta puede poseer y siempre consta de un límite inferior y de un límite superior.

En el ejemplo de la tabla anterior, los números extremos, 50 y 59, son los *límites de clase*; el número menor (50) es el *límite inferior de la clase* y el mayor (59) es el *límite superior de la clase*.

Límites Reales de clase

Considerando la tabla del ejemplo sobre los salarios que se registran con aproximación de dólares, la tercera clase (50 – 59) incluye a todos los salarios desde 49.50 a 59.50 dólares.

La representación exacta de los datos anteriores, 49.5 y 59.5, se denomina *límites reales de clase* o *límites verdaderos de clase*; el menor de ellos (49.5) se identifica como *límite real inferior* y el mayor de ellos (59.5) se identifica como *límite real superior*.

Límite real inferior se determina sumando el límite inferior más el límite superior de la clase contigua anterior y dividiendo entre dos.

Límite real superior se determina sumando el límite superior más el límite inferior de la clase contigua siguiente y dividiendo entre dos.

Ejemplo:

Considerando los intervalos de clase de la tabla de los salarios diarios del ejemplo anterior:

Salarios (dólares)	
Límites inferiores	Límites superiores
*	29
30	39
40	49
50	59
60	69
70	79
80	89
90	99
100	*

Intervalos de clases

Al calcular los límites reales de clases para el primer intervalo de clase resulta:

$$\boldsymbol{LRi = \frac{30 + 29}{2} = 29.5}$$

$$\boldsymbol{LRs = \frac{39 + 40}{2} = 39.5}$$

Para el segundo intervalo de la clase resulta:

$$\boldsymbol{LRi = \frac{40 + 39}{2} = 39.5}$$

$$\boldsymbol{LRs = \frac{49 + 50}{2} = 49.5}$$

Y así sucesivamente, se calculan los límites reales en cada clase.

Por lo anterior, se concluye que los límites reales de clases para la tabla del ejemplo son:

Salarios (dólares)	
Límites reales inferiores	Límites reales superiores
29.5	39.5
39.5	49.5
49.5	59.5
59.5	69.5
69.5	79.5
79.5	89.5
89.5	99.5

Límites reales de clases

La tabla anterior resulta ser incierta, ya que los límites reales de clases no son coincidentes son las observaciones reales; por ejemplo, si una observación fuese 49.5, no es posible definir si pertenece al intervalo de clase (39.5 – 49.5) o al intervalo (49.5 – 59.5). Lo anterior da lugar a que en algunas ocasiones los límites reales de clases sean utilizados únicamente como símbolos de las clases.

1.5.1.4 Marca de clase

Se define como el punto medio de un intervalo de clase y se representa por (X). La marca de clase también se denomina *punto medio de la clase*, y matemáticamente se determina por la suma de los límites inferior y superior del intervalo de clase, dividida entre dos. Lo anterior se simboliza por la ecuación:

$$\boldsymbol{Marca\ de\ clase\ (X) = \frac{L\acute{i}mite\ inferior\ (L_i) + l\acute{i}mite\ superior\ (L_s)}{2}}$$

Ejemplo:

Al considerar los intervalos de clases de la tabla de los salarios diarios del ejemplo anterior, tenemos que la marca de clase del primer, tercer y quinto intervalo, respectivamente, es:

$$\boldsymbol{X_1 = \frac{L_i + L_s}{2} = \frac{30 + 39}{2} = \frac{69}{2} = 34.5}$$

$$\boldsymbol{X_3 = \frac{L_i + L_s}{2} = \frac{50 + 59}{2} = \frac{109}{2} = 54.5}$$

$$\boldsymbol{X_5 = \frac{L_i + L_s}{2} = \frac{70 + 79}{2} = \frac{149}{2} = 74.5}$$

Por lo anterior, se concluye que la marca de clase para la tabla del ejemplo es:

Salarios (dólares)		Marca de clase (X)
Límites reales inferiores	Límites reales superiores	
30 —	39	34.5
40 —	49	44.5
50 —	59	54.5
60 —	69	64.5
70 —	79	74.5
80 —	89	84.5
90 —	99	94.5

En todos los análisis estadísticos, se supone que el total de los elementos pertenecientes a un intervalo de clase son coincidentes con el valor de la marca de clase; por ejemplo, todos los salarios del segundo intervalo de clase (40 – 49) se consideran de 44.5 dólares.

1.5.1.5 Amplitud o ancho de clase

Representa la diferencia que existe entre los límites inferiores o superiores de cada intervalo de clase. Se calcula dividiendo el rango entre el número de clases; matemáticamente se expresa con la ecuación:

$$C = \frac{R}{NC}$$

Se define también como la diferencia entre los límites reales de clase que forman el intervalo de clase y se expresa matemáticamente:

$$C = LRS - LRI$$

La *anchura común* se presenta cuando todos los intervalos de clase de una distribución de frecuencias tienen igual tamaño o longitud, la cual se simboliza por *C* y matemáticamente es igual a la diferencia entre dos límites reales de clase inferiores o superiores sucesivos:

$$\left.\begin{aligned} C &= 39.5 - 29.5 = 10 \\ C &= 49.5 - 39.5 = 10 \\ C &= 59.5 - 49.5 = 10 \end{aligned}\right\}$$ Anchura común

Por lo anterior, se concluye que los intervalos de clase de dicha tabla presentan una anchura común de 10 unidades.

Reglas generales para formar las distribuciones de frecuencias:

1. Determinar el recorrido o rango entre los datos registrados.
2. Dividir el rango en un número razonable de intervalos de clase que tengan el mismo tamaño o anchura. Si lo anterior no es posible, se aconseja emplear intervalos de clase de diferente tamaño o anchura. Por lo general, el número de intervalos de clase se selecciona entre 5 y 20, dependiendo de la cantidad de datos registrados.
3. Determinar las frecuencias de clase, es decir, contar el número de observaciones que caen dentro de cada intervalo de clase.

Ejemplos:

1. En la siguiente tabla se presentan los pesos de 50 estudiantes que se registraron con aproximación de una libra. Construye la tabla de distribución de frecuencias.

128	154	140	122	154	135	159	167	142	144
136	148	130	137	146	143	162	154	146	147
158	116	179	141	139	129	164	175	149	128
136	163	132	137	145	144	150	145	170	181
151	135	125	132	160	156	155	138	153	147

Solución:

a) Al ordenar los datos en forma creciente tenemos:

Dato menor	16	29	35	38	43	46	49	54	59	67	
	22	30	36	39	44	46	50	54	60	70	
	25	32	36	40	44	47	51	55	62	75	
	28	32	37	41	45	47	53	56	63	79	
	28	35	37	42	45	48	54	58	64	**81**	Dato mayor

b) Su rango es:

$$\boldsymbol{R = X_M - X_m}$$

$$\boldsymbol{R = 181 - 116}$$

$$\boldsymbol{R = 65}$$

c) Selección del número de intervalo de clase por medio de la ecuación:

$$\boldsymbol{C = \frac{R}{NC}}$$

Se elige como número de clase a partir del 5:

$\boldsymbol{C = \frac{65}{5} = 13}$ Este número de clase no me sirve, debo encontrar uno en el que el resultado se pueda redondear al inmediato superior.

$\boldsymbol{C = \frac{65}{6} = 10.8 \approx 11}$ Con este número de clases (6) y con la anchura de (11), podremos construir la tabla de manera que tengamos una anchura común y todos los datos, desde el menor hasta el mayor, estén dentro de los intervalos de clase.

d) Formemos los intervalos de clases:

Iniciamos con el dato menor (116) que sería el límite inferior para el primer intervalo de clase. En el segundo intervalo de clase sumamos la anchura C=11 al límite inferior 116. Y seguimos sumando así sucesivamente hasta que lleguemos a la clase 6. Para estar seguros de que está correcto el agrupamiento, podemos agregar una clase más, solo para comprobar que el límite inferior es mayor al dato mayor (181). Al sumar obtenemos: 171+11= **182.** Como se puede observar, este valor es mayor que 181, por lo tanto la agrupación de datos es correcta.

Ahora solo falta obtener los límites superiores de cada intervalo de clase. Esto se logra restando una unidad al límite inferior contiguo siguiente, o sea, para el primer límite superior: 127-1= 126; para el segundo 138-1= 137, así sucesivamente hasta llegar al último límite superior; como recordamos, agregamos un intervalo más que nos dio como límite inferior **182**, este es el valor que tomaremos para restar la unidad y encontrar el límite superior correspondiente al último intervalo o clase, como se observa en la siguiente tabla:

Intervalos de clase (Peso en libras)	
Límite inferior	Límite superior
116	127-1= **126**
116+11= 127	138-1= **137**
127+11= 138	149-1= **148**
138+11= 149	160-1= **159**
149+11= 160	171-1= **170**
160+11= 171	182-1= **181**

e) Se construye la tabla de distribución de frecuencias:

En la *primera columna* se colocan los intervalos de clase que ya se crearon, como encabezado se anota el tipo de observación que se está analizando, o sea, el peso en libras.

En la *segunda columna*, se colocan las frecuencias de cada intervalo de clase, que en este ejemplo, sería el número de estudiantes que tienen los pesos que se observan de cada intervalo. Así que en la tabla de datos ordenados contamos a los estudiantes que pesan entre 116 y 126 libras (se deben contar también los datos que se repiten). Y así seguimos sumando para encontrar las frecuencias de todos los intervalos de clase y al final sumamos todas las frecuencias (N) para comprobar que correspondan al mismo número de estudiantes que se registraron.

En la *tercera columna* se obtienen las marcas de clase de cada intervalo con la siguiente ecuación:

Peso en libras	No. de estudiantes (f)	Marca de clase (X)
116 - 126	3	121
127 - 137	13	132
138 - 148	15	143
149 - 159	11	154
160 - 170	6	165
171 - 181	3	176
	N= 50	

$$X_1 = \frac{L_i + L_s}{2} = \frac{116 + 126}{2} = \frac{242}{2} = 121$$
$$X_2 = \frac{L_i + L_s}{2} = \frac{127 + 137}{2} = \frac{264}{2} = 132$$
$$X_3 = \frac{L_i + L_s}{2} = \frac{138 + 148}{2} = \frac{286}{2} = 143$$
$$X_4 = \frac{L_i + L_s}{2} = \frac{149 + 159}{2} = \frac{308}{2} = 154$$
$$X_5 = \frac{L_i + L_s}{2} = \frac{160 + 170}{2} = \frac{330}{2} = 165$$
$$X_6 = \frac{L_i + L_s}{2} = \frac{171 + 181}{2} = \frac{352}{2} = 176$$

NOTA: Es posible construir otras tablas de distribución de frecuencias utilizando otros criterios sobre la selección del número de intervalos de clase.

Ejercicios:

Construye la tabla de distribución de frecuencias para los siguientes problemas:

1. Las siguientes son las puntuaciones finales que obtuvieron 48 estudiantes en un examen de química orgánica.

63	73	40	77	65	70	58	75
86	90	55	72	73	56	76	69
94	66	84	53	75	89	42	64
65	47	62	88	48	84	77	89
76	86	75	60	64	67	65	69
33	79	69	76	35	49	96	73

2. Las masas en kilogramos de 50 reses destinadas al rastro de Monterrey, Nuevo León son:

336	392	315	318	321	337	365	380	375	345
319	365	381	329	387	358	370	387	370	356
327	353	310	326	318	382	394	358	400	388
346	326	359	360	305	362	376	343	395	363
376	335	317	302	346	379	324	354	335	348

Actividades de cierre

Después de haber explicado el ejemplo anterior con datos enteros y de realizar los ejercicios en las actividades de desarrollo, reforzaremos el tema de la construcción de tablas de distribución de frecuencias, pero con variables continuas, es decir, con datos que contienen 1, 2 o más decimales.

Ejemplos:

1. Los porcentajes de consumo nacional de la producción de trigo durante los últimos 60 años se registran en la siguiente tabla:

72.7	72.5	71.8	77.9	69.8	73	88.9	78.5	73.2	58.7
67.3	67.6	69.6	76.8	72.6	84.6	91.4	69.9	61.4	62.3
69.3	70.1	77.3	80.4	76.5	66.7	83.9	78.6	69.2	68.5
74.6	68.2	73.9	79.1	78.1	87.4	78.4	57.9	65.3	71.7
66.1	75.3	74.7	79.8	82.3	61.6	84.1	62.4	54.8	74.2
65.6	73.4	71.2	81.6	64.2	79.2	89.7	63.7	59.2	64.9

Construye una distribución de frecuencias de los porcentajes, utilizando intervalos de clase adecuados.

a) Después de ordenar los datos en forma creciente tenemos:

Dato menor	54.8	62.3	65.6	68.5	70.1	72.7	74.6	77.9	79.2	84.1	
	57.9	62.4	66.1	69.2	71.2	73	74.7	78.1	79.8	84.6	
	58.7	63.7	66.7	69.3	71.7	73.2	75.3	78.4	80.4	87.4	
	59.2	64.2	67.3	69.6	71.8	73.4	76.5	78.5	81.6	88.9	
	61.4	64.9	67.6	69.8	72.5	73.9	76.8	78.6	82.3	89.7	
	61.6	65.3	68.2	69.9	72.6	74.2	77.3	79.1	83.9	91.4	Dato mayor

b) Su rango es:

$$R = X_M - X_m$$

$$R = 91.4 - 54.8$$

$$R = 36.6$$

c) Selección del número de intervalo de clase por medio de la ecuación:

$$C = \frac{R}{NC}$$

$C = \frac{36.6}{5} = 7.32$ Este NC no es factible, recordemos que buscamos un resultado en el que se pueda redondear al inmediato superior.

$$C = \frac{36.6}{6} = 6.1 \qquad C = \frac{36.6}{7} = 5.22$$

$C = \frac{36.6}{8} = 4.75 \approx 4.8$ Este es el que seleccionamos, ya que se redondea hacia arriba, si se observa como estamos tratando con datos que tienen un decimal, también el redondeo se hace para que nos quede un solo decimal.

NOTA: También se pudiera considerar otra opción, ya que si tomamos 10 clases quedaría el resultado de la división 3.66, que se redondea a 3.7, esta elección depende del analista de acuerdo a lo que quiere observar con la información.

d) Formemos los intervalos de clases:

Como ya se explicó en el ejemplo anterior, primero formamos el límite inferior de cada intervalo de clase, sumando **4.8** al dato menor que es **54.8** y considerando **8 clases** o intervalos. Se van sumando sucesivamente. Al igual que en el ejemplo anterior, agregamos una clase más 88.4+4.8= **93.2**; verificamos que el resultado es mayor al dato mayor de **91.4,** así que la agrupación está correcta.

Intervalos de clase (Porcentajes de producción de trigo)	
Límite inferior	**Límite superior**
54.8	59.6-0.1= 59.5
54.8+4.8= 59.6	64.4-0.1= 64.3
59.6+4.8= 64.4	69.2-0.1= 69.1
64.4+4.8= 69.2	74.0-0.1= 73.9
69.2+4.8= 74.0	78.8-0.1= 78.7
74.0+4.8= 78.8	83.6-0.1= 83.5
78.8+4.8= 83.6	88.4-0.1= 88.3
83.6+4.80= 88.4	93.2-0.1= 93.1

Continuamos y formaremos el límite superior de cada intervalo de clase. En el ejemplo anterior se restaba la unidad a cada límite inferior porque los datos son números enteros, en este ejemplo restaremos **0.1** a cada límite inferior ya que los datos observados tienen un decimal. Para el último límite superior tomamos el valor del límite inferior que agregamos de más que es **93.2.**

Ahora sí ya tenemos todos los intervalos, cabe aclarar, que en este ejemplo, no nos queda como límite superior el dato mayor **91.4**, lo importante es que este dato esté siempre dentro del último intervalo de clase.

e) Se construye la tabla de distribución de frecuencias:

Primera columna: Anotamos los intervalos de clase que formamos en el inciso anterior, y como encabezado Porcentaje de producción ya que son los datos que se están observando.

Porcentaje de producción	Número de Años (f)	Marca de clase (X)
54.8 - 59.5	4	57.15
59.6 - 64.3	6	61.95
64.4 - 69.1	9	66.75
69.2 - 73.9	16	71.55
74.0 - 78.7	12	76.35
78.8 - 83.5	6	81.15
83.6 - 88.3	4	85.95
88.4 - 93.1	3	90.75
	N= 60	

Segunda columna: En la tabla de datos ordenados del inciso b), contamos el número de datos que pertenecen a cada uno de los intervalos de clase y así obtener la frecuencia absoluta. No se olviden de tomar en cuenta todos los datos, aunque éstos estén varias veces.

Columna tres: Obtenemos por medio de su ecuación, la marca de clase o el punto medio de los intervalos o clases.

$$X_1 = \frac{L_i+L_s}{2} = \frac{54.8+59.5}{2} = \frac{114.3}{2} = 57.15$$

$$X_2 = \frac{L_i+L_s}{2} = \frac{59.6+64.3}{2} = \frac{123.9}{2} = 61.95$$

$$X_3 = \frac{L_i+L_s}{2} = \frac{64.4+69.1}{2} = \frac{133.5}{2} = 66.75$$

$$X_4 = \frac{L_i+L_s}{2} = \frac{69.2+73.9}{2} = \frac{143.1}{2} = 71.55$$

$$X_5 = \frac{L_i+L_s}{2} = \frac{74.0+78.7}{2} = \frac{152.7}{2} = 76.35$$

$$X_6 = \frac{L_i+L_s}{2} = \frac{78.8+83.5}{2} = \frac{162.3}{2} = 81.15$$

$$X_7 = \frac{L_i+L_s}{2} = \frac{83.6+88.3}{2} = \frac{171.9}{2} = 85.95$$

$$X_8 = \frac{L_i+L_s}{2} = \frac{88.4+93.1}{2} = \frac{181.5}{2} = 90.75$$

Anotamos todos los resultados de la marca de clase en su respectiva columna. Cabe aclarar que estos datos no se redondean.

2. La siguiente tabla muestra los diámetros interiores de 40 inyectores para motor diesel producidos por una compañía; éstos están registrados con aproximación de milésimas de pulgada. Construye la tabla de distribución de frecuencias.

0.438	0.441	0.435	0.431	0.426	0.437	0.428	0.424
0.429	0.444	0.44	0.436	0.436	0.439	0.439	0.437
0.443	0.436	0.446	0.433	0.442	0.434	0.435	0.43
0.44	0.434	0.439	0.432	0.437	0.438	0.437	0.442
0.436	0.432	0.434	0.43	0.435	0.433	0.431	0.438

a) Después de ordenar los datos en forma creciente, se tiene:

0.424	0.43	0.433	0.435	0.436	0.437	0.439	0.442
0.426	0.431	0.433	0.435	0.436	0.438	0.439	0.442
0.428	0.431	0.434	0.435	0.437	0.438	0.44	0.443
0.429	0.432	0.434	0.436	0.437	0.438	0.44	0.444
0.43	0.432	0.434	0.436	0.437	0.439	0.441	**0.446**

b) Su rango es:

$$R = X_M - X_m$$
$$R = 0.446 - 0.424$$
$$R = 0.022$$

c) Selección del número de intervalo de clase por medio de la ecuación:

$$C = \frac{R}{NC}$$

$C = \frac{0.022}{5} = 0.0044$ Este NC no es factible, recordemos que buscamos un resultado en el que se pueda redondear al inmediato superior.

$C = \frac{0.022}{6} = 0.0036 \approx 0.004$ Este es que seleccionamos, ya que se redondea hacia arriba, si se observa como estamos tratando con datos que tienen 3 decimales, también el redondeo se hace para que nos queden tres decimales.

NOTA: Recuerden que se puede seleccionar cualquier otra opción en la que se pueda redondear al inmediato superior.

d) Formemos los intervalos de clases:

Como ya se explicó en el ejemplo anterior, primero formamos el límite inferior de cada intervalo de clase, sumando **0.004** al dato menor que es **0.424** y considerando **6 clases** o intervalos. Se van sumando sucesivamente. Al igual que en el ejemplo anterior, agregamos una clase más 0.444+0.004= **0.448**; verificamos que el resultado es mayor al dato mayor de **0.446,** así que la agrupación está correcta.

Intervalos de clase (Porcentajes de producción de trigo)	
Límite inferior	Límite superior
0.424	0.428-0.001= 0.427
0.424+0.004= 0.428	0.432-0.001= 0.431
0.428+0.004= 0.432	0.436-0.001= 0.435
0.432+0.004= 0.436	0.440-0.001= 0.439
0.436+0.004= 0.440	0.444-0.001= 0.443
0.440+0.004= 0.444	0.448-0.001= 0.447

Continuamos y formaremos el límite superior de cada intervalo de clase. En el ejemplo anterior se restaba **0.1** a cada límite inferior ya que los datos observados tienen un decimal. En este ejemplo restaremos **0.001**, porque estamos tratando con datos de 3 decimales.

Para el último límite superior tomamos el valor del límite inferior que agregamos de más que es **0.448.**

Ahora sí, ya tenemos todos los intervalos, cabe aclarar, que en este ejemplo, no nos queda como límite superior el dato mayor **0.446**, lo importante es que este dato esté siempre dentro del último intervalo de clase.

e) Se construye la tabla de distribución de frecuencias:

Primera columna: Anotamos los intervalos de clase que formamos en el inciso anterior, y como encabezado Diámetros en pulgadas ya que son los datos que se están observando.

Segunda columna: En la tabla de datos ordenados del inciso b), contamos el número de datos que pertenecen a cada uno de los intervalos de clase y así obtener la frecuencia absoluta. No se olviden de tomar en cuenta todos los datos, aunque éstos estén varias veces.

Columna tres: Obtenemos por medio de su ecuación, la marca de clase o el punto medio de los intervalos o clases.

Diámetros en pulgadas	**Número inyectores (f)**	**Marca de clase (x)**
0.424 - 0.427	2	0.4255
0.428 - 0.431	6	0.4295
0.432 - 0.435	10	0.4335
0.436 - 0.439	14	0.4375
0.440 - 0.443	6	0.4415
0.444 - 0.447	2	0.4455
	N= 40	

$$X_1 = \frac{L_i+L_s}{2} = \frac{0.424+0.427}{2} = \frac{0.851}{2} = 0.4255 \qquad X_4 = \frac{L_i+L_s}{2} = \frac{0.436+0.439}{2} = \frac{0.875}{2} = 0.4375$$

$$X_2 = \frac{L_i+L_s}{2} = \frac{0.428+0.431}{2} = \frac{0.859}{2} = 0.4295 \qquad X_5 = \frac{L_i+L_s}{2} = \frac{0.440+0.443}{2} = \frac{0.883}{2} = 0.4415$$

$$X_3 = \frac{L_i+L_s}{2} = \frac{0.432+0.435}{2} = \frac{0.867}{2} = 0.4335 \qquad X_6 = \frac{L_i+L_s}{2} = \frac{0.444+0.447}{2} = \frac{0.891}{2} = 0.4455$$

Anotamos todos los resultados de la marca de clase en su respectiva columna. Cabe aclarar que estos datos no se redondean.

Ejercicios:

1. Los siguientes son los diámetros, en pulgadas, de 42 tubos requeridos para perforar en primera fase, igual número de pozos petroleros en el sureste de la República Mexicana.

   Construye la tabla de distribución de frecuencias. La ordenación de los datos deberá ser en forma creciente.

23.3	22	22.6	25.9	25.6	23.8	26.0
24.2	23.5	23.7	21.8	23.8	22.7	25.2
25.4	25.2	24.2	26.3	24.2	22.9	23.4
24.6	25.8	22.3	25.1	25.4	23.6	24.1
26.7	24.4	25.6	24.8	24.9	24.7	25.2
23.6	26.1	24.8	25.4	23.9	22	22.6

2. Los diámetros interiores de 63 arandelas fabricadas por una compañía, que se miden en pulgadas, son:

0.583	0.571	0.574	0.582	0.565	0.574	0.580	0.549	0.548
0.582	0.568	0.573	0.553	0.584	0.568	0.576	0.552	0.572
0.554	0.592	0.593	0.571	0.577	0.546	0.554	0.553	0.586
0.548	0.573	0.581	0.564	0.563	0.568	0.575	0.582	0.576
0.563	0.563	0.593	0.572	0.558	0.589	0.591	0.555	0.544
0.542	0.580	0.564	0.589	0.590	0.593	0.563	0.559	0.560
0.592	0.566	0.552	0.590	0.580	0.584	0.584	0.557	0.586

Construye la tabla de distribución de frecuencias con los datos de la tabla anterior. Ordena los datos en forma ascendente.

Ejercicios Adicionales

Realiza los siguientes ejercicios como reforzamiento de los temas vistos anteriormente.

Instrucciones: Construye la tabla de distribución de frecuencias de los datos que se te proporcionan. Ordena de forma ascendente.

1. Los siguientes datos representan los ingresos en miles de pesos de 56 personas de una cierta compañía.

28	23	19	19	18	23	20	20
27	21	23	17	15	16	19	19
17	31	18	17	25	13	20	17
24	26	22	26	11	19	15	23
24	20	24	21	22	21	19	25
17	23	20	18	27	24	20	29
22	21	12	21	26	16	16	17

2. Los siguientes datos representan las ventas en millones de pesos de 60 compañías en el ramo textil.

35	35	33	24	28	21	37	13	6	23
18	24	28	31	12	14	45	33	28	26
29	22	43	10	43	43	24	19	9	35
42	36	23	33	21	21	29	11	32	14
17	19	21	13	38	23	19	34	25	7
11	9	36	4	30	29	29	18	36	10

3. En la siguiente tabla de datos están representadas las estaturas en metros de 64 niños menores de 7 años.

0.06	0.90	0.08	0.65	0.07	0.17	0.28	0.72
0.78	0.51	0.61	0.86	0.60	0.76	0.53	0.16
0.88	0.68	0.96	0.77	0.41	0.74	0.08	0.67
0.46	0.37	0.47	0.86	0.45	0.06	0.82	0.81
0.55	0.96	0.60	0.59	0.52	0.21	0.41	0.45
0.87	0.99	0.07	0.17	0.45	0.59	0.52	0.61
0.53	0.46	0.45	0.28	0.38	0.35	0.62	0.17
0.16	0.57	0.20	0.82	0.65	0.72	0.99	0.20

1.6 Representación gráfica e interpretación.

Introducción

Uno de los objetivos de la estadística es comunicar los resultados de una investigación de manera clara y concisa, las representaciones gráficas de esos resultados obedecen a esta intención.

Actividades de Apertura

Una forma de comunicar una gran cantidad de temas es usar las gráficas, las vemos a diario para decirnos si va subiendo o bajando el dólar, la gasolina, los contagios por CoVid-19, los índices de obesidad, desempleo, entre otros.

En los medios de comunicación se usan con mayor frecuencia las gráficas que las tablas estadísticas, ya que la mayoría de la población no comprende las tablas y mucho menos lee un texto o artículo que trate de temas que incluya información estadística de utilidad.

¿Sabes cuál es uno de los países con mayor índice de pobreza entre los miembros de OCDE? Otra pista es uno de los primeros lugares en consumo de refresco, otra pista sus habitantes ocupan uno de los primeros lugares en obesidad. Toda esta información casi siempre se muestra en gráficas estadísticas.

Actividades de Desarrollo

Una vez concentrados los datos en una tabla, se puede hacer la representación gráfica correspondiente. Un Gráfico Estadístico es la representación de los datos estadísticos por medio de figuras geométricas (puntos, líneas, rectángulos, etc.) cuyas dimensiones son proporcionales al valor numérico de los datos.

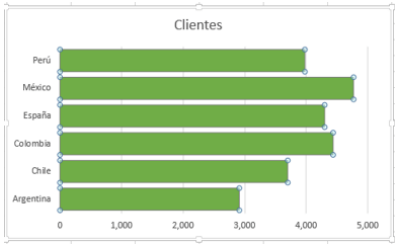
El gráfico es útil para dar una rápida idea de la situación general que se está analizando, permite determinar con un simple vistazo máximos, mínimos, variaciones del fenómeno y determinar soluciones.

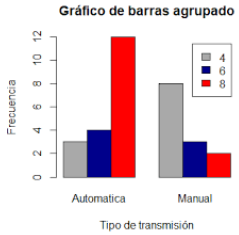
1.6.1. Histograma.

Un histograma es la representación de la información mediante una gráfica en forma de barras, sin espacios entre ellas, esta imagen nos permite apreciar visualmente la distribución de un conjunto de datos. Los histogramas son de utilidad para darnos una idea general de la distribución de una población o muestra.

a) Gráfico de Barras.

La construcción de este gráfico se basa en la representación de un valor numérico por un rectángulo. Las barras son usualmente verticales, pero también pueden ser horizontales o compuestas.



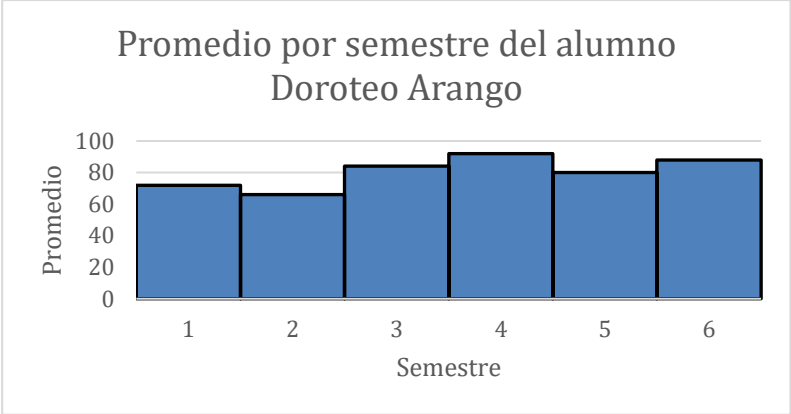


Ejemplo 1.

De la tabla siguiente construye un histograma:

Promedio por semestre del alumno Doroteo Arango

Semestre	Promedio
1	72
2	66
3	84
4	92
5	80
6	88



1.6.1.1 Frecuencias absolutas

Cuando se grafican a mano las frecuencias, es mucho más sencillo graficar las frecuencias con sus valores originales (valores absolutos), cuando los cálculos y graficación se realizan en una hoja de cálculo o software similar, la dificultad es la misma.

1.6.1.2 Frecuencias relativas

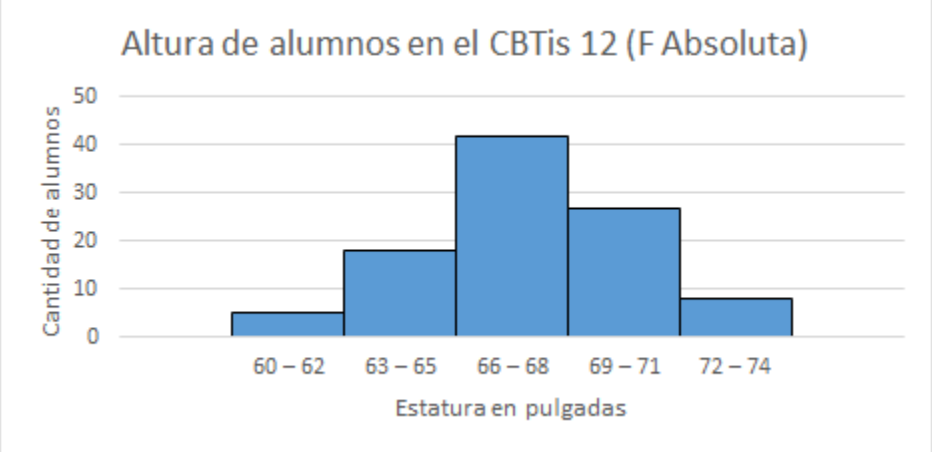
El trabajar con las frecuencias relativas implica el uso de porcentajes y en muchos casos el uso de decimales. Cuando se grafican a mano las frecuencias relativas es menos sencillo que sus valores originales (valores absolutos). La forma de los histogramas es la misma, solo cambian los valores del eje Y.

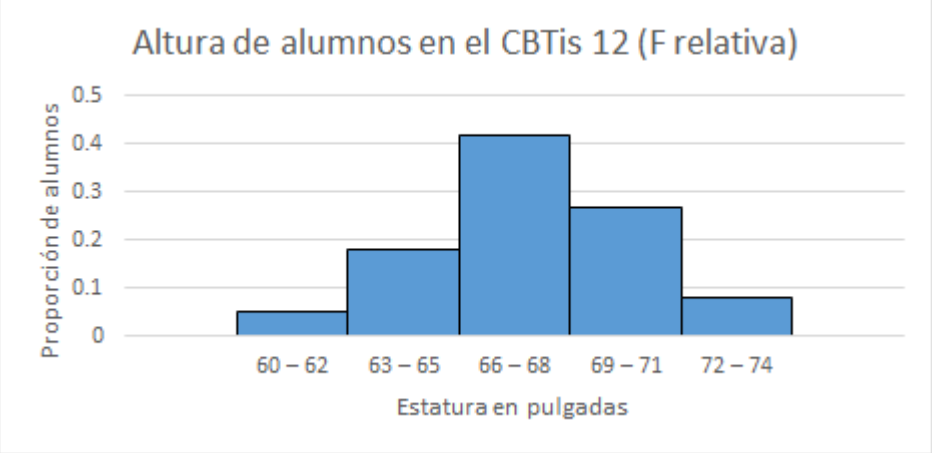
Ejemplo comparativo entre Frecuencias absolutas y relativas.

Altura de 100 estudiantes del CBTIS 12

Altura (pulg)	Cantidad	F relativa	F relativa (%)
60 – 62	5	0.05	5%
63 – 65	18	0.18	18%
66 – 68	42	0.42	42%
69 – 71	27	0.27	27%
72 – 74	8	0.08	8%
	N= 100	**N= 1.00**	**N= 100%**

Las gráficas de la tabla anterior serían, primeramente la de frecuencia absoluta y después la frecuencia relativa.





1.6.2. Polígono de Frecuencias

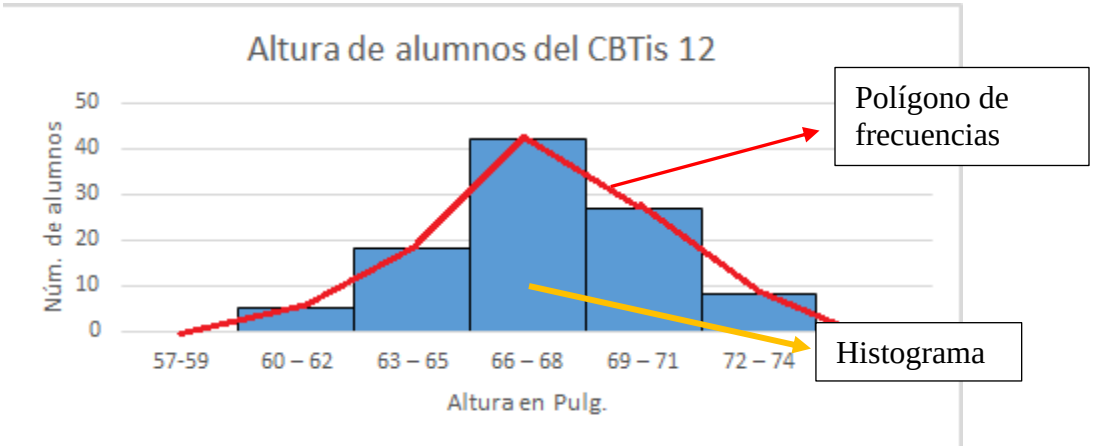
Para complementar el histograma, podemos unir los centros de las barras sobre el eje "x", es decir las marcas de clase, además de prolongar estas marcas a línea inferior y superior. Al conjunto de esta poligonal se le denomina "**polígono de frecuencias".**

Ejemplo 2. De la tabla siguiente construye el histograma.

Altura de 100 estudiantes del CBTIS 12

Altura (pulg)	Cantidad
60 – 62	**5**
63 – 65	**18**
66 – 68	**42**
69 – 71	**27**
72 – 74	**8**

Primeramente, se determinan las **marcas de clase:**
$X_1 = \frac{60+62}{2} = 61$, $X_2 = \frac{63+65}{2} = 64$, $X_3 = \frac{66+68}{2} = 67$, $X_4 = \frac{69+71}{2} = 70$, $X_5 = \frac{72+74}{2} = 73$, también se determina la inferior y superior, que son 58 y 76

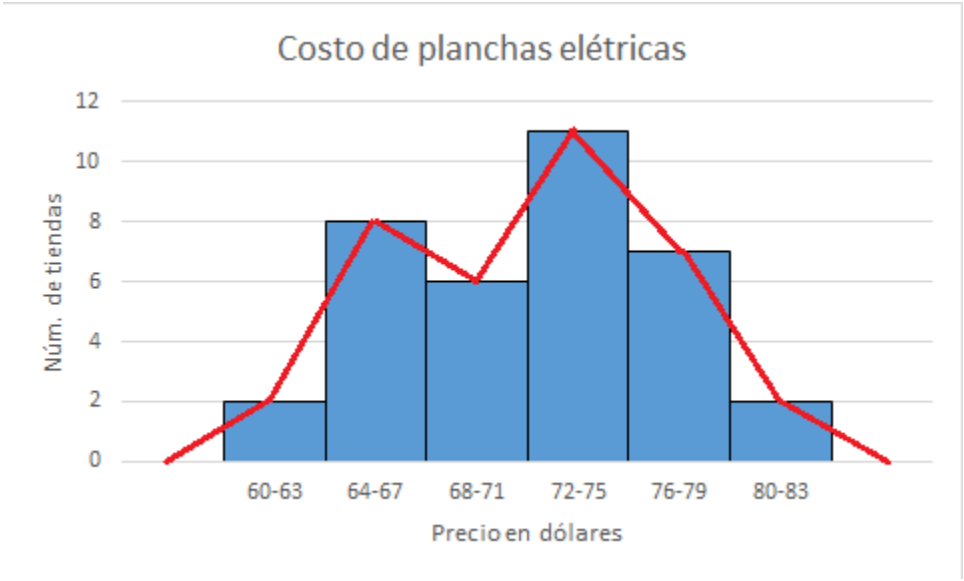


Ejemplo 3. La siguiente tabla muestra el costo de planchas eléctricas en dólares, elabora su histograma con su polígono de frecuencia.

Costo	No. de tiendas	Fa
60 - 63	2	2
64 - 67	8	10
68 - 71	6	16
72 - 75	11	27
76 - 79	7	34
80 - 83	2	36
Suma	36	

Marca de clase:

$X_1 = \frac{60+63}{2} = 61.5$, $X_2 = \frac{64+67}{2} = 65.5$, $X_3 = \frac{68+71}{2} = 69.5$, $X_4 = \frac{72+75}{2} = 73.5$, $X_5 = \frac{76+79}{2} = 77.5$, $X_6 = \frac{80+83}{2} = 81.5$; inferior 57.5, superior 85.5



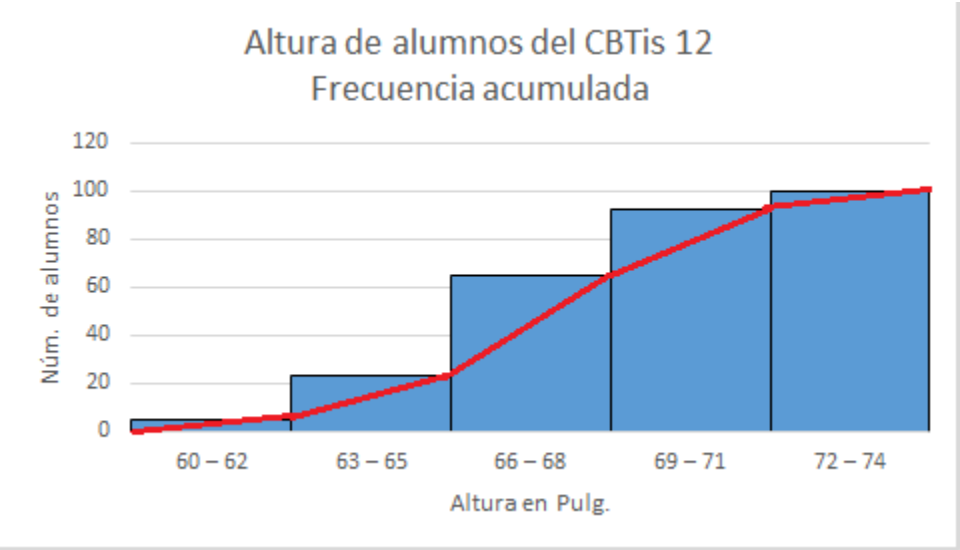
1.6.3. Ojiva.

La ojiva es similar al polígono de frecuencias, solo que en lugar de usar la parte media de las barras en su elaboración, usaremos el vértice inferior izquierdo del primer rectángulo y lo unimos con una línea hasta el vértice superior derecho del primer rectángulo, continuamos la línea hasta el vértice superior derecho del siguiente rectángulo y así sucesivamente. La segunda diferencia es el uso de las frecuencias acumuladas.

Ejemplo 4. De la tabla del ejemplo 2, construye la ojiva usando la frecuencia acumulada.

Altura de 100 estudiantes del CBTIS 12

Altura (pulg)	Cantidad	Fa
60 – 62	5	5
63 – 65	18	23
66 – 68	42	65
69 – 71	27	92
72 – 74	8	100
Suma	**100**	



1.6.4. Gráfico Circular, de sectores o Pastel

Es una forma alternativa del gráfico de barras, en su construcción se utiliza una circunferencia dividida en sectores angulares proporcional al valor de la variable. Es conveniente usar las fra para facilitar su trazo

Ejemplo 1. De la tabla anterior construye un gráfico circular con los promedios.

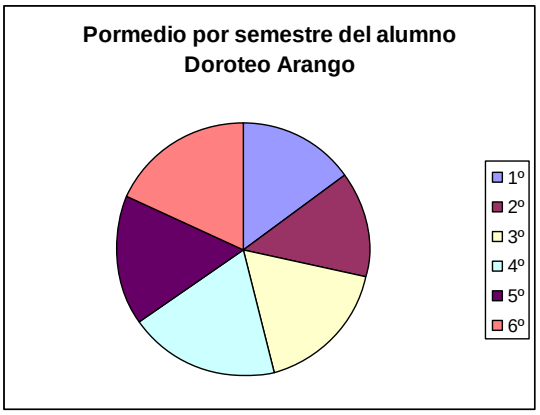
Procedimiento: Utilizaremos como opción el uso de la regla de 3 (razones y proporciones).

$$\frac{360°}{X} = \frac{482}{72} \rightarrow x = 53.77° \cong 54°$$ (Se requiere transportador para trazar)

De igual forma podremos calcular los ángulos para los siguientes promedios.

66 – 49º 84 – 63º 92- 69º 80 – 60º 88 – 65º

Otra opción es utilizar las frecuencias relativas acumuladas y utilizar el 100% como toda la circunferencia.



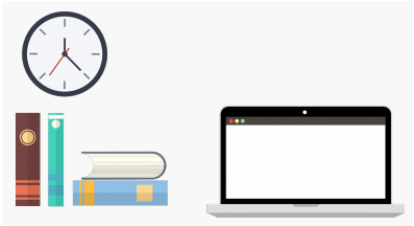
Nota: Si contamos con una hoja electrónica de cálculo (Excel), el programa nos puede hacer cualquier gráfica, pero debemos saber hacerla usando instrumentos de dibujo.

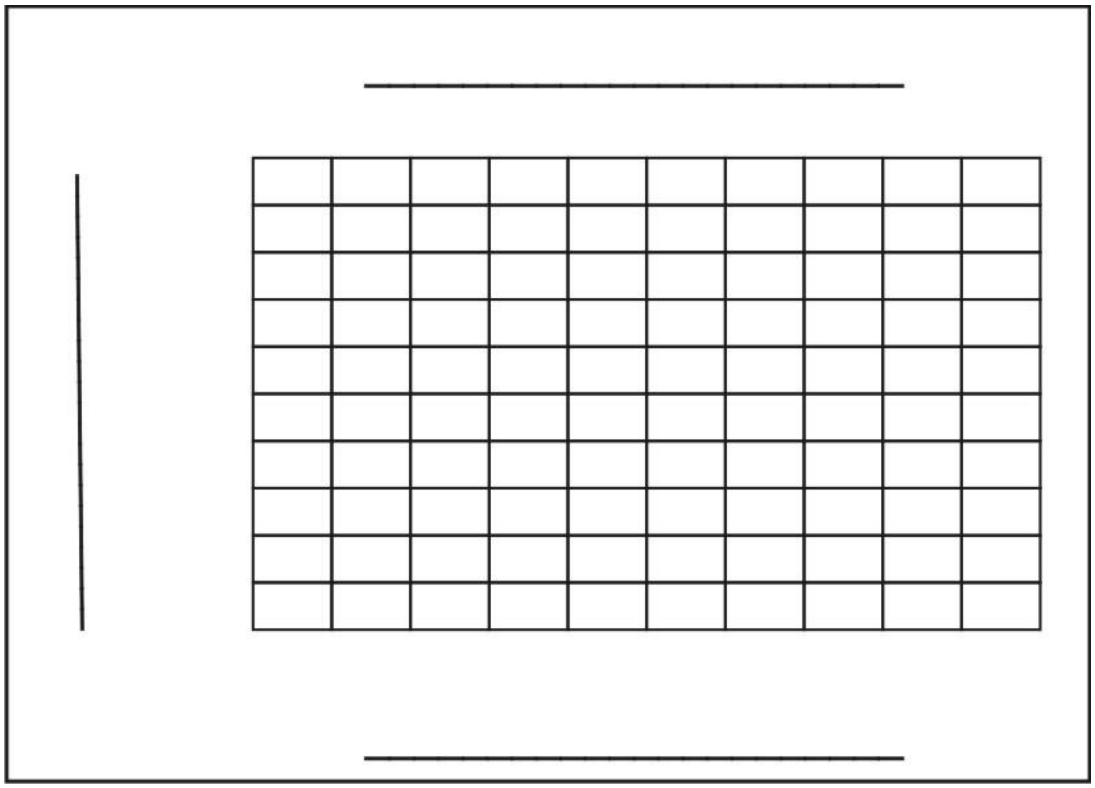
EJERCICIO 1:

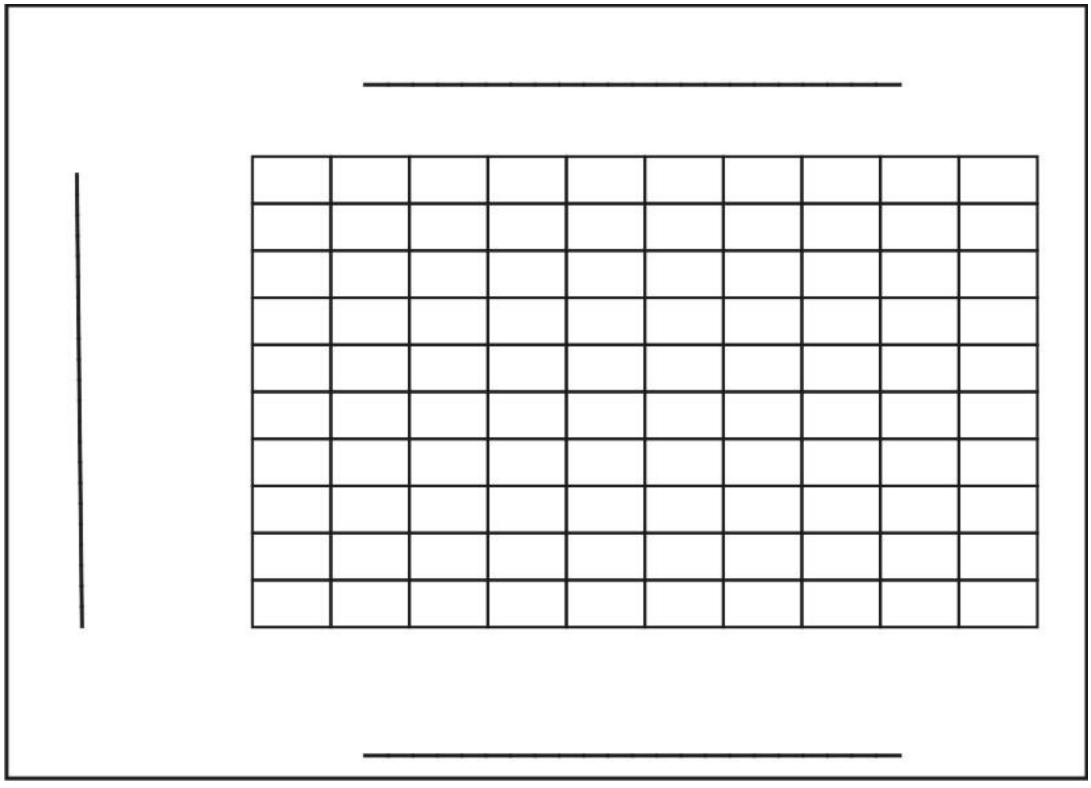
La antigüedad en años de trabajo de un conjunto de docentes es la siguiente, según el departamento de Recursos Humanos del CBTIS 12 en 2006, como se muestra en la siguiente tabla.

Con ellos traza un HISTOGRAMA CON POLÍGONO DE FRECUENCIAS Y EN EL SEGUNDO ESPACIO UNA OJIVA USANDO EL Fa

Años	Cantidad	fr %	fa	fra %
3 – 5	5	20	5	20
6 – 8	6	24	11	44
9 – 11	7	28	18	72
12 – 14	7	28	25	100







Actividades de cierre

Como te pudiste dar cuenta, el trazo de una gráfica usando tu juego de geometría, puede ser una actividad muy bonita, pero laboriosa.

Las gráficas pueden imprimirles calidad y profesionalismo a tus trabajos, de igual forma, pueden ayudarte a argumentar tu postura en una exposición.

Nuevamente, te invitamos a que apliques las herramientas como la hoja de cálculo (Excel), que te permitiría hacer esto en una fracción del tiempo. Es importante que vayas relacionando lo aprendido en cada área del conocimiento con lo que ya conoces.

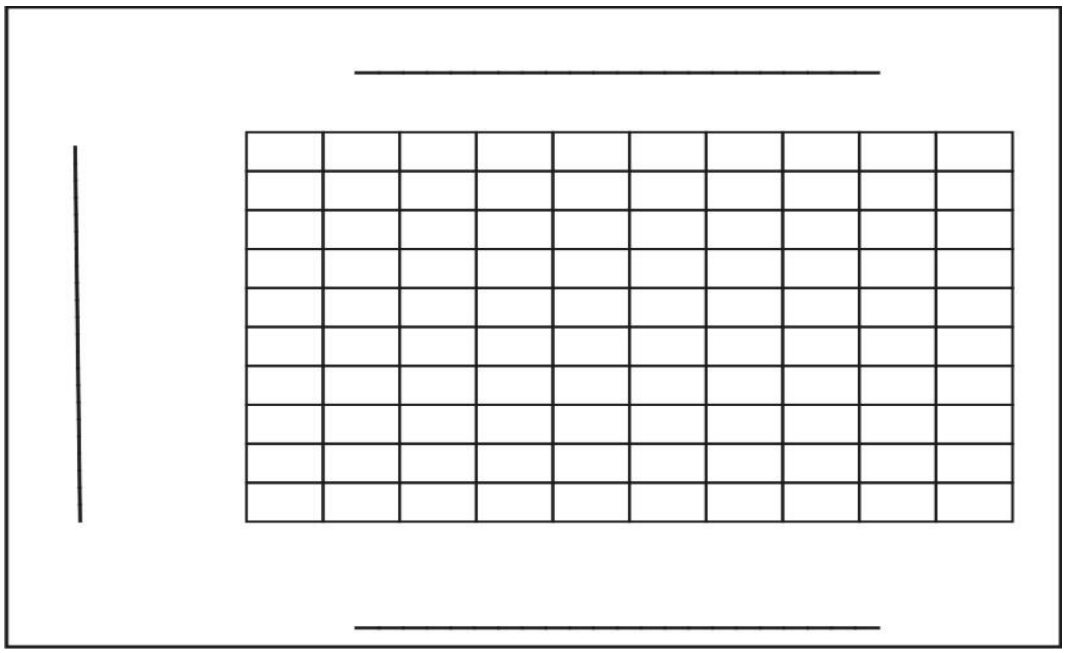
¡Date la oportunidad de ejercitar tus habilidades digitales!

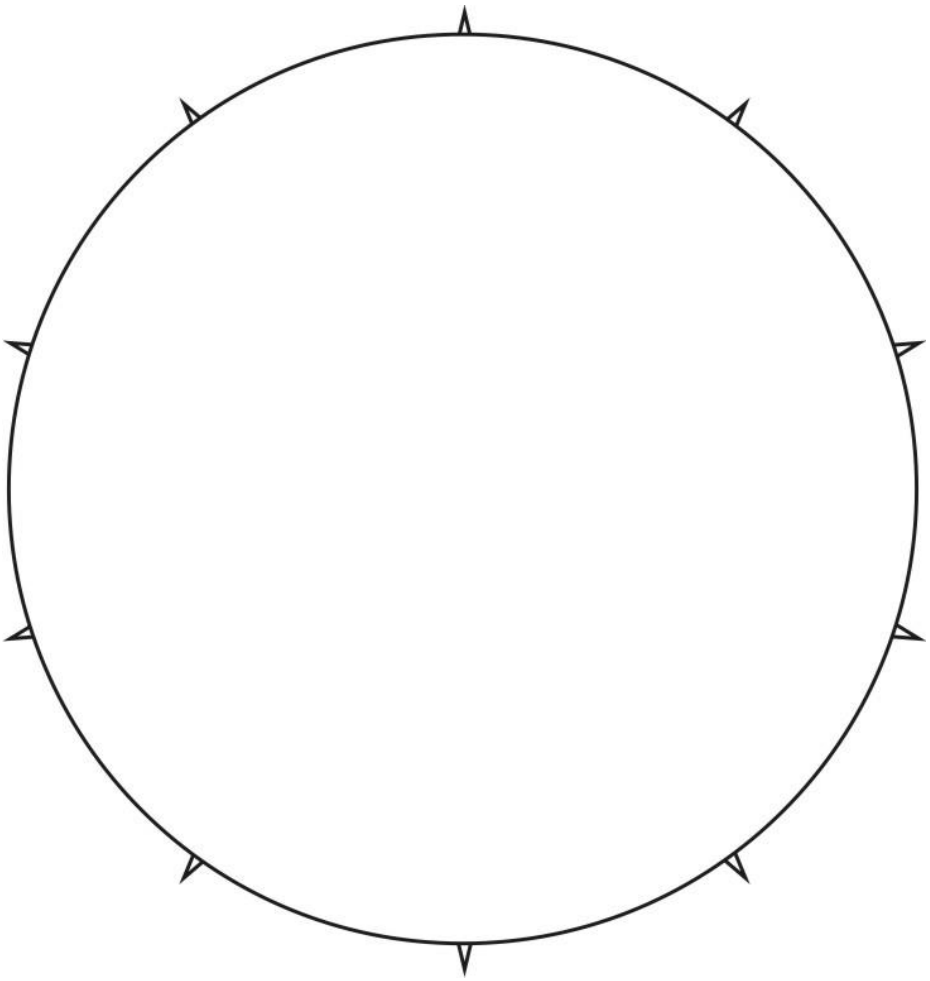
Ejercicios Adicionales

EJERCICIO 2: Con la siguiente información contenida en la tabla estadística, construye un HISTOGRAMA Y UNA GRÁFICA CIRCULAR (se recomienda usar el fra)

Las películas exhibidas en el DF según su nacionalidad en 1988 sin contar las de EUA son: mexicanas 705, alemanas 75, francesas 176, inglesas 71, italianas 142, otras 399, fuente Anuario Estadístico 1990.

Nacionalidad	Películas	fr %	fa	fra %
Mexicana	705	44.96%	705	**44.96%**
Alemana	75	4.78%	780	**49.74%**
Francesa	176	11.22%	956	**60.97%**
Inglesa	71	4.53%	1027	**65.50%**
Italiana	142	9.06%	1169	**74.55%**
Otras	**399**	**25.45%**	**1568**	**100.00%**





Bloque 2 | Medidas de tendencia central

2.1 Media

Introducción

La tendencia central se refiere al punto medio de una distribución. Las medidas de tendencia central se conocen también como medidas de posición. Las tres medidas de tendencia central o de centralización más importantes son la Media Aritmética o media, Mediana y Moda.

Media

Un valor que es simbólico y representativo de un conjunto de datos se denomina promedio. Se hace notar que en un conjunto de datos ordenados de acuerdo con su magnitud, el promedio siempre tiende a situarse en el centro de dicho conjunto, razón por la cual los promedios se denominan también medidas de centralización o de tendencia central.

2.1.1. Media aritmética

Es el promedio más utilizado y por lo general se denomina media. La media aritmética o media de un conjunto de elementos se define como la suma de los valores de estos elementos dividido entre el número total de ellos. Se simboliza $\bar{X}$ (léase $\bar{X}$ barra), para **datos no agrupados**, matemáticamente se determina por la ecuación:

$$\bar{X} = \frac{\sum X_i}{N} = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N}$$

Donde:
$\bar{X}$= Media aritmética o media.
X= Valor de cada elemento.
N= Número total de elementos.

Ejemplo:
Determina la media aritmética de los números 18, 13, 15, 22 y 20.

$$\bar{X} = \frac{18 + 13 + 15 + 22 + 20}{5} = \frac{88}{5} = \mathbf{17.6}$$

Media aritmética para datos agrupados.

Cuando los datos se presentan agrupados mediante una distribución de frecuencias, todos los valores caen dentro de los intervalos de clase dados que se consideran coincidentes con las marcas de clase o puntos medios de cada intervalo, es decir, las ponderaciones son las frecuencias y las marcas de clase o puntos medios son los valores que se ponderan.

En una distribución de frecuencias para datos agrupados, la media aritmética se determina al multiplicar las distintas marcas o puntos medios de clase por sus respectivas frecuencias de clase; se suman los productos y el resultado se divide entre el número total de frecuencias. Matemáticamente se representa:

$$\bar{X} = \frac{\sum f\, X}{N}$$

Donde:

$\bar{X}$= Media aritmética o media.
X= Marcas o puntos medios de clase.
f= Frecuencia total de clase.
N= Número total de frecuencias.

Ejemplo:

La siguiente tabla de distribución de frecuencias representan las edades de los 96 asistentes a un curso de informática.

Intervalos (edades)	Frecuencias (asistentes)	Marca de clase (X)	$f\, X$
24 - 28	26	8	208
29 - 33	31	17	527
34 - 38	36	29	1 044
39 - 43	11	23	943
44 - 48	46	15	690
49 - 53	51	4	204
	N= 96		$\sum f\, X$ = 3 616

Al sustituir los datos anteriores en la ecuación para la media aritmética resulta:

$$\bar{X} = \frac{\sum f\, X}{N} = \frac{3\,616}{96} = 37.666$$

La media aritmética de las edades es **37.666 años**.

2.2 Mediana y Moda

2.2.1 Mediana

Se define como el **valor medio** a aquel que divide a un conjunto de datos que se ordena de acuerdo con su magnitud (de forma ascendente o de forma descendente) en dos partes iguales, es decir, es aquel valor central que deja por debajo igual número de elementos que por arriba de él.

Si el conjunto de elementos ordenados de acuerdo con su magnitud, es **impar**, la mediana será el valor intermedio de dicha sucesión. Si el conjunto de elementos ordenados de acuerdo con su magnitud es **par**, la mediana será la media aritmética de los dos elementos medios.

Ejemplos:

1. Determina la mediana de los números 25, 29, 24, 29, 27, 25, 29, 31, 26

Solución:
Al ordenar ascendentemente el conjunto de elementos numéricos dado tenemos:

24,25, 25, 26, **27**, 29, 29, 29, 31

Por definición, la mediana de los números es el valor central, en este caso es **27**.

2. Determina la mediana de los números 49, 38, 46, 40, 36, 42, 36, 43

Solución:
Al ordenar descendentemente el conjunto de elementos numéricos dado se tiene:

49, 46, 43, **42, 40**, 38, 36, 36

La mediana es la media aritmética de los números 42 y 40, es decir:

$$Med = \frac{42 + 40}{2} = 41$$

La mediana del conjunto de elementos numéricos dados es **41**

Mediana para datos agrupados.

Para determinar la mediana en datos agrupados se emplea el método de la interpolación, es decir, se fundamenta de la suposición de que los elementos en la **clase mediana** (es la clase en la cual se localiza el valor de la mediana) están distribuidos uniformemente en todo el intervalo.

El cálculo de la mediana de una distribución de frecuencias se determina matemáticamente por la ecuación:

$$\boldsymbol{Me = LR_1 + \left[\frac{\frac{N}{2} - \sum f_1}{f\ med}\right] C}$$

Donde:

LR_1= Límite real inferior de la clase mediana, es decir, la clase que contiene el valor de la mediana.

N= Número total de datos, es decir, la frecuencia total.

$\sum f_1$= Suma de las frecuencias de todas las clases por debajo de la clase mediana.

$f\ med$= Frecuencia de la clase mediana, la cual nunca debe ser mayor que $\frac{N}{2}$

C= El tamaño del intervalo de la clase mediana.

Ejemplo:

La siguiente tabla de distribución de frecuencias muestra las puntuaciones que obtuvieron durante el primer parcial 56 estudiantes en un examen de geometría analítica. Calcula la mediana.

	Intervalos (puntuaciones)	Frecuencias (estudiantes)	
	45 - 51	1	
	52 - 58	3	
	59 - 65	7	
	66 - 72	12	
Clase mediana →	**73 - 79**	15	← *fmed*
	80 - 86	9	
	87 - 93	6	
	94 - 100	3	
		N= 56	

Solución:

Se localiza la clase mediana: $\frac{N}{2} = \frac{56}{2} = 28$

La suma de las frecuencias de las cuatro y cinco primeras clases es 1 + 3 + 7 + 12 = 23 y 1 + 3 + 7 + 12 + 15 =38 , respectivamente, por lo que es evidente que la mediana se encuentra en la quinta clase, qué será, por lo tanto, la clase mediana.

Entonces observamos los siguientes datos:

$$LR_1 = \frac{73 + 72}{2} = 72.5$$

$N = 56$

$\sum f_1 = 23$

$f\ med = 15$

$C = 7$

Al sustituir los datos anteriores en la ecuación de la mediana, resulta:

$$Me = LR_1 + \left[\frac{\frac{N}{2} - \sum f_1}{f\ med}\right] C = 72.5 + \left[\frac{\frac{56}{2} - 23}{15}\right](7)$$

$$Me = 72.5 + \left[\frac{28\ - 23}{15}\right](7) = 72.5 + \left(\frac{5}{15}\right)(7)$$

$$Me = 72.5 + 2.33 = 74.83\ \approx\ 75$$

La media de las puntuaciones calculada a partir de datos agrupados es 75

2.2.2 Moda

Se define como el valor que se presenta con la **mayor frecuencia**, es decir, es el **valor más común** de un conjunto de elementos numéricos dado.

Entre las características de la moda destacan que ésta puede no existir, incluso si existe puede no ser única. Si un conjunto de valores dado presenta una sola moda se denomina **unimodal**. Sí presenta dos modas se denomina **bimodal**; si se presentan más de dos modas se denomina **multimodal**.

Ejemplo 1: Dado el siguiente conjunto de números 4, 6, 8, 10, 12, 14, 16, 18, 20; determina su moda.

Solución: Analizando el conjunto de números dados, se observa que todos ellos tienen igual frecuencia, por lo que se concluye que no presenta moda, es decir, **no existe**.

Ejemplo 2: Dado el siguiente conjunto de números 3, 3, 4, 6, 8, 8, 8, 11, 11, 13, 15; determina su moda.

Solución: Analizar el conjunto de números dado, se observa que el número que se presenta con mayor frecuencia es el 8, por lo que se concluye que el valor de la **moda es 8** y el conjunto se denomina **unimodal**.

Ejemplo 3: Dado el siguiente conjunto de números 10, 12, 12, 12, 13, 14, 14, 14, 15, 15, 16, 17, 17, 17, 19; determina su moda.

Solución: Al analizar el conjunto de números dado, se observa que los números que se presentan con mayor frecuencia son el 12, 14 y 17, por lo que se concluye que los valores de la moda son **12, 14 y 17**, en donde el conjunto se denomina **multimodal**.

Moda para datos agrupados.

Para determinar la moda en datos agrupados se emplea el método de la interpolación, es decir, se fundamenta en encontrar primero la clase modal (es la clase en la cual se localiza el valor de la moda) que se caracteriza por tener la máxima frecuencia de la distribución.

El siguiente paso es determinar la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase **premodal** o anterior; así mismo, la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase **posmodal** o siguiente. Lo anterior es con el fin de observar hacia dónde se concentra el valor modal.

El cálculo de la moda de una distribución de frecuencias se determina matemáticamente por la ecuación:

$$\boldsymbol{Mo = LR_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) C}$$

Donde:

LR_1= Límite real inferior de la clase modal, es decir, la clase que contiene el valor de la moda.

Δ_1= Es la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase premodal o anterior.

Δ_2= Es la diferencia absoluta entre la frecuencia de la clase modal y la frecuencia de la clase posmodal o siguiente.

C = Tamaño del intervalo de clase modal

Ejemplo 1:

La siguiente tabla de distribución de frecuencias agrupa los salarios semanales en miles de pesos de 130 empleados técnicos de la empresa TRW, división de cinturones de seguridad; determina la moda de dichos salarios.

Solución: Primero, se considera la tabla de distribución de frecuencias para localizar la clase modal que se caracteriza por tener la máxima frecuencia, y el valor de la moda que resulta ser (750 – 773).

Intervalos (salarios)	Frecuencias (empleados)	
654 - 677	3	
678 - 701	7	
702 - 725	13	
726 - 749	19	← $\Delta_1 = 26 - 19 = 7$
750 - 773	26	← (Clase modal →)
774 - 797	21	← $\Delta_2 = 26 - 21 = 5$
798 - 821	17	
822 - 845	12	
846 - 869	8	
870 - 893	4	
	N= 130	

También se observan los siguientes datos:

LR_1= El límite real inferior de la clase modal se obtiene sumando el límite superior de un intervalo de clase y el límite inferior de intervalo de clase contiguo superior y dividido entre dos, es decir:

$$LR_1 = \frac{750 + 749}{2} 749.5$$

$LR_1 = 749.5$
$\Delta_1 = 26 - 19 = 7$
$\Delta_2 = 26 - 21 = 5$
$C = 23$

Al sustituir los datos anteriores en la ecuación de la moda resulta:

$$\boldsymbol{Mo = LR_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) C = 749.5 + \left(\frac{7}{7+5}\right)(23)}$$

$$Mo = 749.5 + 13.42 = \mathbf{762.92}$$

La moda de los salarios calculada a partir de datos agrupados es **762. 92 miles de pesos**.

Actividades de Apertura

1. La siguiente tabla de distribución de frecuencias muestra el número de horas a la semana que 60 familias miran programas televisivos; determina la media aritmética.

Intervalos (horas)	Frecuencias (familias)	Marca de clase X	f_x
30 - 39	4		
40 - 49	9		
50 - 59	13		
60 - 69	19		
70 - 79	12		
80 - 89	3		
	N= 60		

$$\bar{X} = \frac{\sum f\, X}{N} =$$

Actividades de Desarrollo

1. Con los datos de la tabla de distribución de frecuencias anterior, calcula:

a) La mediana por fórmula
b) La moda
c) Constuye el Histograma

Intervalos (horas)	Frecuencias (familias)
30 - 39	4
40 - 49	9
50 - 59	13
60 - 69	19
70 - 79	12
80 - 89	3
	$N=$ 60

Actividades de cierre

1. Realiza la siguiente actividad en una hoja de cálculo Excel para determinar lo que se te solicita:

La siguiente tabla de Distribución de frecuencias registra el número de horas al mes que 180 miembros de un club altruista dedican a labores sociales en su comunidad.

Determinar:

a) La media aritmética
b) La mediana
c) La moda
d) Construye el Histograma

Intervalos (horas)	Frecuencias (miembros) f	Marca de clase X	f X
10 - 14	6		
15 - 19	14		
20 - 24	23		
25 - 29	28		
30 - 34	19		
35 - 39	11		
40 - 44	7		
	N= 108		$\sum f\,X =$

Solución:

a) $\bar{X} =$ _____$horas$ b) $M_e =$ _____$horas$ c) $M_o =$ _____$horas$

Ejercicios Adicionales

Como reforzamiento de los temas vistos, realiza los siguientes ejercicios:

I. Resuelve los siguientes problemas para **Datos No Agrupados**.

1. Determina la media aritmética, la mediana y la moda para los siguientes conjuntos de datos numéricos. Menciona también qué tipo de moda resulta en cada uno (unimodal, bimodal o multimodal).

   a) 125, 116, 119, 121, 121, 124, 129
   b) 8, 12, 16, 8, 10, 12, 10, 14, 10, 12, 10, 8, 10

2. El costo de un litro de gasolina *magna sin* en diferentes municipios de la franja fronteriza de la República Mexicana, registrado en el año 2019 es: $15.99, $11.62, $13.50, $15.48, $12.00, $13.50. Determina:

   a) La media aritmética de los costes.
   b) La mediana de los costes.
   c) La moda de los costes.
   d) Indica el tipo de moda.

3. Doce vendedores de autos usados vendieron en una semana, respectivamente, 59, 72, 86, 45, 64, 48, 69, 90, 43, 95, 64 y 81 automóviles. Determina:

   a) La media aritmética.
   b) La mediana.
   c) La moda e indica el tipo de moda que resulta.

II. Resuelve los siguientes problemas para **Datos Agrupados**.

1. La tabla de distribución de frecuencias registra la cantidad de adultos en una población a partir de los 30 años, arrojando los siguientes resultados:

INTERVALO	FRECUENCIA *f*	MARCA DE CLASE X	*f X*
30 - 39	7		
40 - 49	12		
50 - 59	19		
60 - 69	16		
70 - 79	10		
80 - 89	6		
90 - 99	2		
	N = 72		$\sum f\,X =$

Calcula los valores de:

a) La media aritmética
b) La mediana
c) La moda
d) Construye el Histograma

2. En la siguiente tabla de distribución de frecuencias se muestran las estaturas de un grupo de 202 estudiantes. Calcule la media mediana, moda y construye el gráfico de histograma en una hoja de cálculo Excel.

INTERVALO (estaturas)	FRECUENCIA *f*	MARCA DE CLASE X	*f X*
152 - 156	25		
157 - 161	18		
162 - 166	36		
167 - 171	55		
172 - 176	53		
177 - 181	15		
	N = 202		

Soluciones:

a) $\bar{X} =$ $de\ estatura$

b) $M_e =$ $de\ estatura$

c) $M_o =$ $de\ estatura$

2.3 Cuartiles, Deciles y Percentiles

Introducción

Dado un conjunto de elementos que se ordenan de acuerdo con su magnitud, el valor medio que divide al conjunto de datos en dos partes iguales o la media aritmética de los valores medios de dicho conjunto, dan como resultado a la mediana.

Asociadas a la definición de mediana, tenemos otras medidas que se fundamentan en las divisiones proporcionales que pueden hacerse en datos agrupados o sin agrupar y que se denominan cuantiles. Existen diferentes tipos: **Cuartiles, Deciles y Percentiles.**

2.3.1. Cuartiles

Se definen como los intervalos dentro de los cuales quedan proporcionalmente repartidos los datos sin agrupar o agrupados en una distribución formada por 4 partes iguales.

Se tienen 3 cuartiles que se simbolizan por Q_1 (primer cuartil), Q_2 (segundo cuartil) y Q_3 (tercer cuartil), en donde cada uno contendrá el mismo número de datos, es decir, el 25% del total.

Para determinar el valor de los cuartiles en **datos no agrupados**, aplicamos la expresión general $\frac{KN}{4}$ y las siguientes reglas:

a) Si *KN* es divisible exactamente 4, el valor del cuartil buscado será la media aritmética entre el dato ordenado cuya posición se obtiene de $\frac{KN}{4}$ y el siguiente dato de orden.

b) Si *KN* no es divisible exactamente entre 4, el valor del cuartil buscado será el dato ordenado cuya posición si obtiene de redondear al entero más próximo el resultado de $\frac{KN}{4}$.

El cálculo de los cuartiles de una **distribución de frecuencias o datos agrupados** se determina matemáticamente por la ecuación general:

$$Q_K = LR_1 + \left[\frac{\frac{KN}{4} - \sum f}{f_k}\right] C$$

Donde:

K= *K*-enésimo cuartil = Se refiere al primero, segundo y tercer cuartil, según sea el que se tenga que determinar.

LR_1= Límite real inferior de la clase *K*-enésimo cuartil.

N= Número total de datos, es decir, la frecuencia total.

$\sum f$= Suma de las frecuencias de todas las clases por debajo de la clase de *K*-enésimo cuartil.

f_k= Frecuencia de la clase del *K*-enésimo cuartil, la cual nunca debe ser mayor que $\frac{KN}{4}$

C= Tamaño del intervalo de la clase del *K*-enésimo cuartil.

2.3.2 Deciles

Se definen como los intervalos dentro de los cuales quedan proporcionalmente repartidos los datos sin agrupar o agrupados en una distribución formada por 10 partes iguales. Se tienen nueve deciles que se simbolizan por D_1 (primer decil), D_2 (segundo decil), D_3 (tercer decil), …, D_9 (noveno decil), en donde cada uno contendrá el mismo número de datos, es decir, el 10% del total.

Para determinar el valor de los deciles en **datos no agrupados**, aplicamos la expresión general $\frac{KN}{10}$.

El cálculo de los deciles de una **distribución de frecuencias o datos agrupados** se determina matemáticamente por la ecuación general:

$$D_K = LR_1 + \left[\frac{\frac{KN}{10} - \sum f}{f_k}\right] C$$

Donde:

K= *K*-enésimo decil = Se refiere al primero, segundo, tercero,…, noveno decil, según sea el que se tenga que determinar.

LR_1= Límite real inferior de la clase *K*-enésimo decil.

N= Número total de datos, es decir, la frecuencia total.

$\sum f$= Suma de las frecuencias de todas las clases por debajo de la clase de *K*-enésimo decil.

f_k= Frecuencia de la clase del *K*-enésimo decil, la cual nunca debe ser mayor que $\frac{KN}{10}$

C= Tamaño del intervalo de la clase del *K*-enésimo decil.

2.3.3 Percentiles

Se define como los intervalos dentro de los cuales quedan proporcionalmente repartidos los datos sin agrupar o agrupados en una distribución formada por cien partes iguales.

Se tienen noventa y nueve percentiles que se simbolizan por P_1 (primer percentil), P_2 (segundo percentil), P_3 (tercer percentil),…, P_{99} (noventa y nueve percentil), en donde cada uno contendrá el mismo número de datos, es decir, el 1% del total.

Para determinar el valor de los percentiles en **datos no agrupados**, aplicamos la expresión general $\frac{KN}{100}$.

El cálculo de los percentiles de una distribución de frecuencias o datos agrupados se determina matemáticamente por la ecuación general:

$$\boldsymbol{P_K = LR_1 + \left[\frac{\frac{KN}{100} - \sum f}{f_k}\right] C}$$

Donde:

K= *K*-enésimo percentil = Se refiere al primero, segundo, tercero,…, nonogésimo noveno percentil, según sea el que se tenga que determinar.

LR_1= Límite real inferior de la clase *K*-enésimo percentil.

N= Número total de datos, es decir, la frecuencia total.

$\sum f$= Suma de las frecuencias de todas las clases por debajo de la clase de *K*-enésimo percentil.

f_k= Frecuencia de la clase del *K*-enésimo percentil, la cual nunca debe ser mayor que $\frac{KN}{100}$

C= Tamaño del intervalo de la clase del *K*-enésimo percentil.

Ejemplo para datos no agrupados:

Para el siguiente conjunto de datos ordenados:

76, 77, 81, 82, 83, 88, 92, 95, 99, 102, 105, 107, 108, 110, 112, 116, 119, 123, 125, 128, 131, 135, 137, 140, 142, 146, 149, 152, 155, 158, 161, 164, 166, 168, 171.

Determina el Q_3, D_6 y P_{70}.

Solución:

Como el conjunto tiene $N = 35$ datos:

a) $Q_3 = \frac{KN}{4} = \frac{3(35)}{4} = 26.25 \approx 26$ redondeado al entero más próximo.

b) $D_6 = \frac{KN}{10} = \frac{6(35)}{10} = 21$ el valor del D_6 será la media aritmética entre el 21° dato ordenado y el siguiente del conjunto, es decir,

$D_6 = \frac{131+135}{2} = 133$ El valor del D_6 es 133.

c) $P_{70} = \frac{KN}{100} = \frac{70(35)}{100} = 24.5 \approx 25$ redondeado al entero más próximo.

Ejemplo para datos agrupados:

La siguiente tabla de distribución de frecuencias registra los pesos en libras de 42 atletas mexicanos que participaron en los *XXIX Juegos Olímpicos* en Beijing en 2008.

Encuentra:

a) Cuartil Q_1
b) Decil D_9
c) Percentil P_{65}

Intervalos (libras)	Frecuencias (atletas)
110 - 116	2
117 - 123	4
124 - 130	7
131 - 137	9
138 - 144	13
145 - 151	6
152 - 158	1
	$N=$ 42

Solución:

a) Para determinar el primer cuartil Q_1, usamos los siguientes datos:

Datos	Formula	Sustitución
$\frac{KN}{4}=\frac{1(42)}{4}=10.5$	$Q_1 = LR_1 + \left[\frac{\frac{KN}{4}-\sum f}{f_k}\right]C$	$Q_1 = 123.5 + \left[\frac{10.5-6}{7}\right](7)$

$$Q_1 = 123.5 + 4.5 = 128$$

$LR_1 = \frac{123+124}{2} = 123.5$

$\sum f = 2+4 = 6$

$f_k = 7$

$C = 7$

El valor del primer cuartil es 128 libras, es decir, 25% de los atletas pesa 128 libras o menos.

b) Para determinar el noveno decil D_9, usamos los siguientes datos:

Datos	Formula	Sustitución
$\frac{KN}{10}=\frac{9(42)}{10}=37.8$	$D_9 = LR_1 + \left[\frac{\frac{KN}{10}-\sum f}{f_k}\right]C$	$D_9 = 144.5 + \left[\frac{37.8-35}{6}\right](7)$

$$D_9 = 144.5 + 3.27 = 147.77$$

$LR_1 = \frac{144+145}{2} = 144.5$

$\sum f = 2+4+7+9+13 = 35$

$f_k = 6$

$C = 7$

El valor del noveno decil es 147.77 libras, es decir, 90% de los atletas pesa 147.77 libras o menos.

c) Para determinar el sexagésimo quinto percentil P_{65}, usamos los siguientes datos:

Datos	Formula	Sustitución
$\frac{KN}{100} = \frac{65(42)}{100} = 27.3$	$P_{65} = LR_1 + \left[\frac{\frac{KN}{100} - \sum f}{f_k}\right] C$	$P_{65} = 137.5 + \left[\frac{27.3 - 22}{13}\right](7)$
		$P_{65} = 137.5 + 2.85 = 140.35$
$LR_1 = \frac{137 + 138}{2} = 137.5$		
$\sum f = 2 + 4 + 7 + 9 = 22$		
$f_k = 13$		
$C = 7$		

El valor del sexagésimo quinto percentil es 140.35 libras, es decir, 65% de los atletas pesa 143.75 libras o menos.

Actividades de Apertura

1. La siguiente tabla de distribución de frecuencias registra los tiempos en segundos que 35 atletas varones tardan en recorrer los 100 metros planos. Determina:

a) Los cuartiles: Q_1 y Q_2;
b) Los deciles: D_3, D_6 y D_9
c) Los percentiles P_{35}, P_{50}, P_{65} y P_{80}.

Intervalos (segundos)	Frecuencias (atletas)
8.8 - 9.2	2
9.3 - 9.7	5
9.8 - 10.2	13
10.3 - 10.7	9
10.8 - 11.2	6
	N= 35

Actividades de Desarrollo

1. Con los datos calculados de la tabla de distribución de frecuencias anterior:

Intervalos (segundos)	Frecuencias (atletas)	Marca de clase X
8.8 - 9.2	2	
9.3 - 9.7	5	
9.8 - 10.2	13	
10.3 - 10.7	9	
10.8 - 11.2	6	
	N= 35	

a) Construye el Histograma porcentual indicando en la gráfica el valor de los cuantiles: Q_1, D_6 y P_{80}

b) Constuye el Polígono de frecuencias porcentual indicando en la gráfica el valor de los cuantiles: Q_2, D_9 y P_{65}

Actividades de cierre

1. La siguiente tabla de distribución de frecuencias registra las calificaciones de un segundo examen parcial de Física II de un grupo de la Facultad de Ingeniería Mecánica y Eléctrica.

Determina:

a) El cuartil Q_3
b) Los deciles D_4 y D_7
c) Los percentiles P_{25}, P_{50} y P_{90}

Intervalos (calificaciones)	Frecuencias (estudiantes)	Marca de clase X
30 - 39	1	
40 - 49	2	
50 - 59	4	
60 - 69	7	
70 - 79	11	
80 - 89	8	
90 - 100	3	
	N= 33	

Indica también:

d) La calificación más baja lograda por el 20% superior de la clase.
e) La calificación más alta lograda por el 25% inferior de la clase.
f) ¿Cuál es la calificación del 50% de los alumnos?

Actividades de contexto o Transversales

Realiza la siguiente actividad en una hoja de cálculo Excel:

Los siguientes datos son el resultado de una encuesta realizada vía Facebook a 260 estudiantes para conocer el tiempo que emplean diariamente para realizar las tareas que sus maestros les encargan en línea por estos momentos de pandemia.

12	12	4	12	2	5	7	5	7	3	4	2	2	3
5	7	6	6	5	12	10	12	2	12	2	3	4	6
3	4	8	10	12	13	3	4	5	7	5	8	7	3
5	10	11	6	9	3	4	6	8	5	5	4	5	4
5	6	8	6	6	7	3	4	2	5	5	9	2	3
4	6	2	4	5	2	5	5	7	3	4	6	9	1
1	2	10	11	5	6	5	6	5	7	8	6	6	10
1	2	5	6	6	8	3	11	2	5	6	7	2	4
2	7	3	4	6	8	5	8	5	8	4	6	2	5
5	8	12	2	6	5	7	5	7	9	10	2	4	5
7	2	3	5	7	4	6	6	2	4	6	4	5	5
7	5	8	2	5	2	3	2	5	6	7	2	3	5
8	4	6	8	14	8	2	4	5	8	3	6	12	4
6	2	6	0	3	6	4	7	2	4	4	7	6	8
2	3	2	3	3	4	5	6	2	5	2	3	6	9
4	5	3	5	2	3	3	6	6	3	5	4	6	4
8	3	6	2	4	4	7	4	5	5	4	10	3	5
3	4	8	2	4	5	7	4	5	3	5	3	6	4
5	4	7	2	5	2	3	4						

A partir de ellos calcula en una hoja de excel:

1. Rango, Clases y Amplitud que te permitan…
2. Elaborar una tabla de distribución de frecuencias que contenga: Li, Ls, Marca de clase (X), frecuencia relativa (f), frecuencia relativa acumulada (fa), Frecuencia absoluta (F), Frecuencia absoluta acumulada (Fa), Porcentaje (%), Ojiva >, Ojiva <, Grados para la elaboración de la gráfica circular y el producto de la marca de clase por la frecuencia (Xf) para que te ayuden al cálculo de la media aritmética.
3. Realiza el cálculo de las medidas de tendencia central, Media ($\bar{x}$), Mediana (Me) y Moda (Mo).
4. Calcula los cuantiles: Cuartiles (Q_3), Decil (D_7) y Percentil (P_{30})
5. Realiza la gráfica de: Ojivas, Circular, Histograma, Polígono de frecuencias

Ejercicios Adicionales

Como reforzamiento de los temas vistos, realiza los siguientes ejercicios:

I. Resuelve los siguientes problemas de Cuantiles para **Datos No Agrupados**.

1. Treinta fallas de energía eléctrica duraron 23, 130, 49, 100, 36, 31, 85, 54, 123, 67, 44, 38, 93, 17, 97, 75, 42, 81, 62, 128, 19, 39, 96, 26, 80, 53, 85, 77, 29 y 86 minutos. Determina:

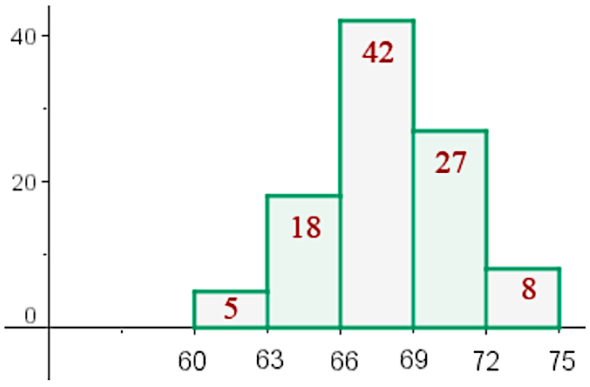
   a) Los cuartiles $Q_1,\ Q_2$ y Q_3
   b) Los deciles $D_1,\ D_3,\ D_5, D_7$ y D_9
   c) Los percentiles $P_5,\ P_{15},\ P_{45},\ P_{65},\ P_{85}$ y P_{95}

2. Quince traductores de inglés-español lograron traducir 92, 107, 123, 90, 78, 81, 76, 94, 105, 88, 109, 121, 95, 101 y 89 páginas de un mismo libro de control de calidad en una hora. Determina:

   a) Los cuartiles Q_1 y Q_3
   b) Los deciles $D_2,\ D_4,\ D_6$ y D_8
   c) Los percentiles $P_{10},\ P_{30},\ P_{50},\ P_{70},\ P_{90}$ y P_{99}

II. Resuelve los siguientes problemas de Cuantiles para **Datos Agrupados**.

2. El histograma de la distribución correspondiente al peso de 100 alumnos de Bachillerato es el siguiente:



   a) A partir de la gráfica anterior, construye la tabla de frecuencias.
   b) ¿A partir de que valores se encuentran el 25% de los alumnos más pesados?

Intervalos (pesos)	Frecuencias (estudiantes)	Marca de clase

2. A partir de los datos de la siguiente tabla de distribución de frecuencias registra las puntuaciones de una prueba de coordinación física dadas a 50 conductores de vehículos que habían consumido alcohol.

INTERVALO (puntuaciones)	FRECUENCIA *f* (conductores)	MARCA DE CLASE X
50 - 59	8	
60 - 69	18	
70 - 79	12	
80 - 89	7	
90 - 99	5	
	N = 50	

Construye:

a) El histograma porcentual indicando el valor de los cantiles: Q_3, D_7 y P_{85}

b) El polígono de frecuencias porcentual indicando el valor de los cantiles: Q_1, D_5 y P_{70}

Bloque 3 | Medidas de dispersión

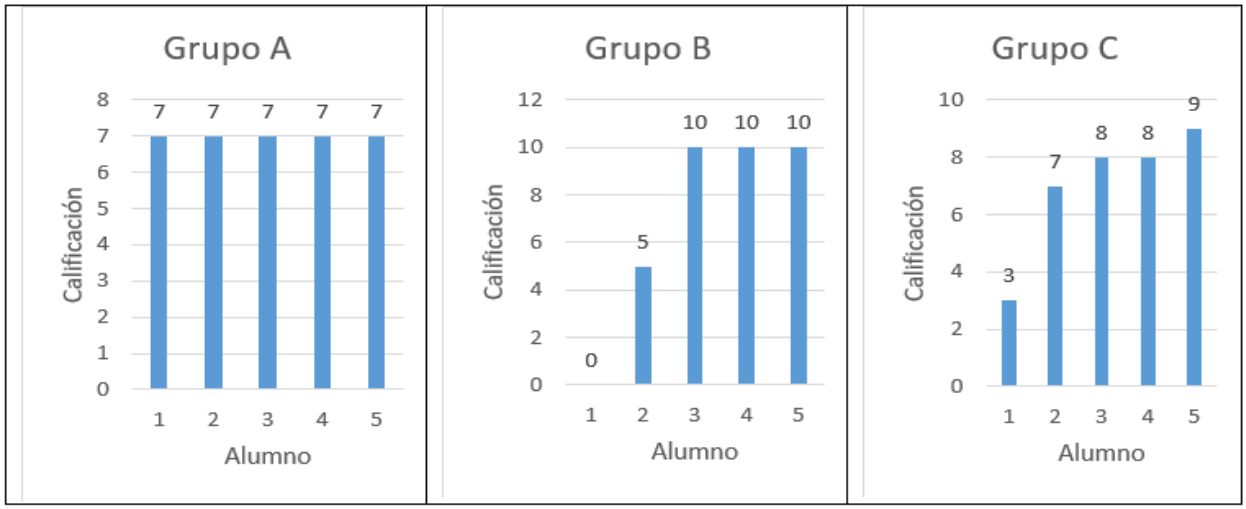
3.1 Rango

Introducción

¿Por qué estudiar la dispersión?

Las medidas de tendencia central, como la media o la mediana, buscan resumir con un solo número, un conjunto de valores. Desde este punto de vista resultan valiosas, pero no dicen nada sobre la dispersión de los datos, es decir, no describen en qué medida difieren entre sí los valores.

Veamos este ejemplo: Se analizan las calificaciones de los estudiantes de tres grupos en un examen de Probabilidad:



Analiza las gráficas anteriores y contesta lo siguiente:

1. ¿Qué similitudes y diferencias observas entre las gráficas?

2. ¿Cuál gráfica presenta mayor dispersión de los datos?, ¿cuál menor dispersión de los datos?, ¿por qué?

3. ¿Existe alguna relación entre la dispersión de los datos (mayor o menor) y la forma de la gráfica? Argumenta tu respuesta.

4. Para comparar los tres grupos, ¿qué medida de tendencia central recomendarías usar?

5. Con base en la información proporcionada, ¿puedes decir cuál de los grupos logró un mejor aprendizaje? Justifica tu respuesta.

Rango

Supongamos que un guía de turismo puede mencionar que una laguna tiene en promedio 1.2 metros de profundidad. Supongamos que usted no sabe nadar, ¿desearía cruzarla sin solicitar mayor información? Quizá no. Probablemente se preguntaría ¿cuántos metros es su máxima profundidad? ¿Cuántos metros es su mínima profundidad?

Si mide 1.4 metros de máxima profundidad y 1.0 metros de mínima profundidad ¿La cruzaría?

Si mide 1.9 metros de máxima profundidad y 0.5 metros de mínima profundidad ¿La cruzaría?

Cuando los datos se encuentran poco dispersos, es decir, acumulados alrededor de la media aritmética, este valor se considera representativa de los datos. Por el contrario, una medida grande de dispersión indica que la media no es confiable.

El *rango o recorrido* es un valor numérico que indica la diferencia entre el valor máximo y el mínimo de una población o muestra estadística. Suele ser utilizado para obtener la dispersión total. Es decir, si tenemos una muestra con dos observaciones: 10 y 100 pesos, el rango será de 90 pesos.

Sobre todo, en finanzas, el rango es muy útil para observar cuán grande podría llegar a ser una variación o cambio. Vale la pena mencionar también que, en no pocas ocasiones, el rango no es una medida fija. Por ejemplo, imaginemos que el crecimiento del Producto Interno Bruto (PIB) de un país, ha estado entre el 3 y el 5% durante los últimos 20 años. El rango para estos datos, será del 2% pero esto no quiere decir que siempre vaya a ser ese. De modo que si en el año 21, el crecimiento es del -1%, el rango de los últimos 21 años, pasará del 2% al 6%.

La fórmula que se emplea es:

$$R = x_M - x_m$$

Dónde:
R = Rango o Recorrido
x_M = Dato mayor de la muestra
x_m = Dato menor de la muestra

Actividades de Apertura

En binas calcular la media aritmética y la diferencia entre la calificación mayor y la menor de cada muestra de grupos en Probabilidad y Estadística:

Grupo A 6, 3, 10, 2 y 9
Grupo B 6, 6, 6, 6 y 6
Grupo C 6, 4, 6, 8 y 6
Grupo D 4, 8, 4, 8 y 6

Grupo	Media	Rango
A		
B		
C		
D		

Actividades de Desarrollo

Contestar las siguientes preguntas:

¿Qué grupo presenta mayor rango en sus calificaciones? _____

¿Qué grupo presenta menor rango en sus calificaciones? _____

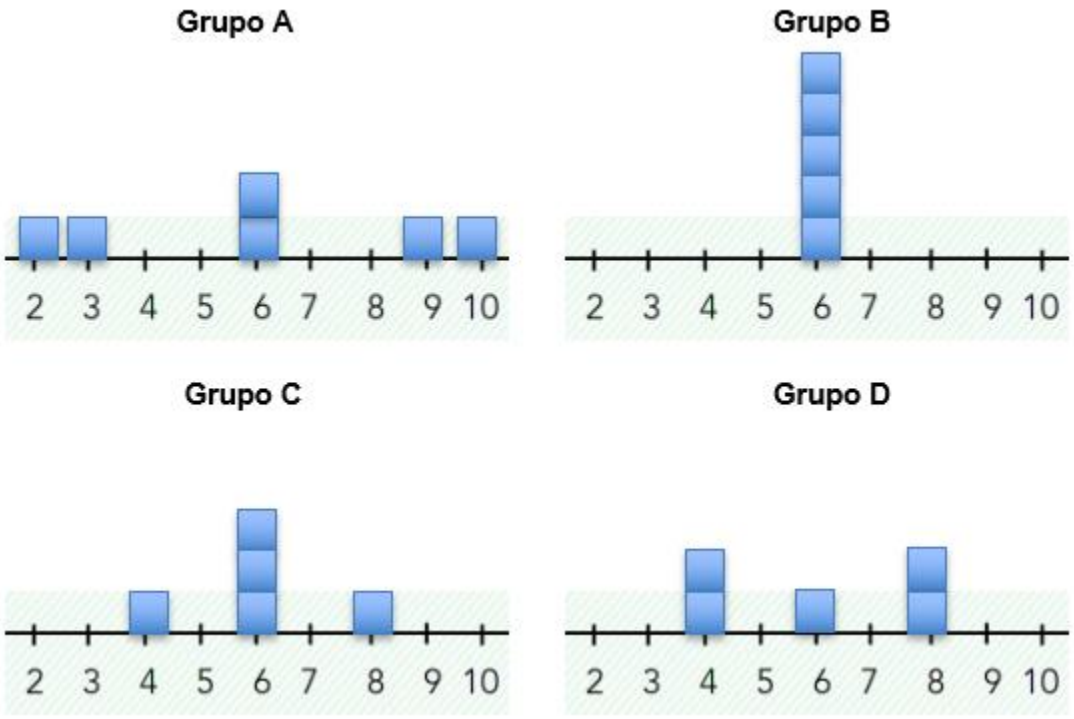
Observa los datos de los grupos C y D. ¿Cuál de ellos presenta mayor variabilidad? _____

Sin embargo, tienen el mismo rango ¿verdad?

Actividades de cierre

Imagina que las siguientes rectas numéricas con balanza en donde el número 6, que es la media aritmética es el punto de equilibrio. Coloca las calificaciones como si fueran bloques cuadrados sobre la recta. El grupo A es el ejemplo.

Imagina que las siguientes rectas numéricas con balanza en donde el número 6, que es la media aritmética es el punto de equilibrio. Coloca las calificaciones como si fueran bloques cuadrados sobre la recta. El grupo A es el ejemplo.



Anota y discute tus observaciones con el resto del grupo.

Actividades de contexto o Transversales

A 5 compañeros de tu grupo hazles las siguientes preguntas y anota en el cuadro las respuestas.

Compañeros	**Estatura en cm**	**Tiempo que tarda en llegar de su casa a la escuela en min.**	**Número de miembros que habitan en su casa**	**Edad en años**
1				
2				
3				
4				
5				
Rango				

Con los datos y el resultado de los rangos obtenidos, discutan en el grupo sus observaciones y escriban sus conclusiones.

Ejercicios Adicionales

1. Los siguientes datos son los clientes diarios de la florería “El Tulipán Negro”: 34, 45, 23, 34, 26, 32, 31, 41. Calcular el Rango.

2. Los miembros de una cooperativa de viviendas tienen las siguientes edades:

42	60	60	38	60	63	21	66	56	57	51	57	44	45	35
30	35	47	53	49	50	49	38	45	28	41	47	42	53	32
54	38	40	63	48	33	35	61	47	41	55	53	27	20	21
42	21	39	39	34	45	39	28	54	33	35	43	48	48	27
53	30	29	53	38	52	54	27	27	43	28	63	41	23	58
56	59	60	40	24										

¿Cuál es la diferencia de edades entre el mayor y el menor de los miembros de la cooperativa? Use la fórmula de Rango.

3.2 Desviación media

Introducción

Esta medida de dispersión nos representa la diferencia absoluta promedio que existe entre cada dato que se encuentra en la muestra y la media de los datos y se determina de la siguiente manera:

$$DM = \frac{\sum_1^n |x_i - \bar{x}|}{n}$$

Dónde:

x_i = dato i (o primer dato)

$\bar{x}$= media aritmética de la muestra

n = número de datos en la muestra

La interpretación de los resultados se interpreta como el grado de *alejamiento absoluto promedio* de los datos con respecto a su media.

¿Por qué sacar el valor absoluto de las diferencias entre cada dato y la Media Aritmética? Si solo se hicieran diferencias entre cada dato y la media aritmética, éstas tendrían signos positivos y negativos ya que algunos datos son menores que la media y otros son mayores que la media, luego al sumar las diferencias, con sus signos correspondientes, éstas se irían anulando unas con otras, dando como resultado cero y no sería posible medir el grado de alejamiento promedio de los datos en la muestra.

Actividades de Apertura

Considerando las calificaciones de Probabilidad y Estadística de los alumnos de los grupos del ejemplo anterior. Calcular, en cada caso, las sumas de las desviaciones. Recuerde que todos los grupos tienen media 6. El grupo A es el ejemplo:

Para el grupo A: $\sum_{i=1}^{5}(x_i - \bar{x}) = (6-6) + (3-6) + (10-6) + (2-6) + (9-6) =$

Para el grupo B:

Para el grupo C:

Para el grupo D:

Actividades de Desarrollo

Como se puede observar en las actividades de apertura, la suma de las desviaciones siempre es cero. Esto se debe a que, como la media es el valor central, la suma de las diferencias positivas es el inverso aditivo de la suma de las diferencias negativas (si esto no pasara, la balanza no se equilibrara).

1. Ahora, usando el valor absoluto de las desviaciones, calcula la Desviación Media de cada grupo:

Para el grupo A: $DM = \frac{\sum_{i=1}^{5}|x_i - \bar{x}|}{n} = \frac{|6-6|+|3-6|+|10-6|+|2-6|+|9-6|}{5} =$

Para el grupo B:

Para el grupo C:

Para el grupo D:

¿Qué podrías poner como observaciones? _____

2. Pregunta a 5 de tus compañeros la calificación que obtuvieron en el primer parcial de Probabilidad y estadística y anótalas en la siguiente tabla y complétala:

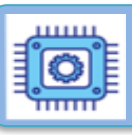
Compañero	**Calificación**	**Media $\bar{x}$**	$\sum_{i=1}^{5}(x_i - \bar{x})$	$DM = \frac{\sum_{i=1}^{5}\lvert x_i - \bar{x}\rvert}{n}$
1				
2				
3				
4				
5				

Actividades de cierre

Contesta las siguientes preguntas:

1. ¿Puede servir la Desviación Media para medir la variabilidad total de un conjunto de datos? _____

2. ¿Por qué? _____



Actividades de contexto o Transversales

Pregunta a 5 compañeros diferentes lo siguiente, y de cada conjunto de datos calcular la Desviación Media.

1. ¿Cuántos mensajes de WhatsApp has recibido en el transcurso del día?

2. ¿En este momento cuántos chavos(as) te gustan?

3. ¿Cuántos km son los que más has viajado en cualquier medio de transporte?

Ejercicios Adicionales

Resuelve los siguientes ejercicios:

1. Una muestra de personas arrojaron los siguientes datos que corresponden al número de hermanos que tienen: 3, 0, 1, 0, 2, 0, 1. Calcular la desviación de estos datos.

3. Hallar la desviación media de los siguientes datos (segundos que tardaron algunos alumnos en contestar una pregunta): 3, 5, 8, 2, 10, 6, 3, 4, 4 y 5.

3.3 Varianza y desviación típica o estándar

Introducción

La varianza es una medida de dispersión que representa la variabilidad de una serie de datos respecto a su media. Formalmente se calcula como la suma de las diferencias al cuadrado divididos entre el total de observaciones.

¿Por qué se elevan al cuadrado los residuos?

La razón por la que las diferencias se elevan al cuadrado es sencilla. Si no se elevasen al cuadrado, la suma de diferencias sería cero. Así pues, para evitarlo, tal como ocurre con la desviación típica se elevan al cuadrado. El resultado de la varianza puede reflejar datos absurdos como litros cuadrados, pesos cuadrados, etc. Para que tenga sentido la interpretación calcularíamos la desviación típica que es la raíz cuadrada de la varianza.

La desviación típica o estándar se hace para poder trabajar en las unidades de medida iniciales. Claro que, como es normal, uno puede preguntarse, ¿de qué sirve tener como concepto la varianza? Bien, aunque la interpretación del valor que arroja no nos da demasiada información, su cálculo es necesario para obtener el valor de otros parámetros.

Para calcular la covarianza, que la veremos en temas posteriores, necesitamos la varianza y no la desviación típica, para calcular algunas matrices econométricas se utiliza la varianza y no la desviación típica. Es una cuestión de comodidad a la hora de trabajar con los datos, según lo que deseemos calcular. En lo general, la desviación estándar es la raíz cuadrada de la varianza.

También se puede calcular como la desviación estándar o típica al cuadrado. Se usan las siguientes fórmulas:

3.3.1 Varianza

Varianza, para datos NO agrupados

Para una muestra:

$$\mathcal{S}^2 = \frac{\sum_{i=1}^{n}(m_i - \bar{x})^2}{n-1}$$

Para una población:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(m_i - \mu)^2}{N}$$

Dónde:

x_i = es el dato i en la posición i

$\bar{x}$= media aritmética de la muestra

μ= media aritmética de la población

n = número de datos de la muestra

N = número de datos de la población

Varianza, para datos agrupados

Para una muestra:

$$s^2 = \frac{\sum_{i=1}^{n} f_i\left(m_i - \bar{x}\right)^2}{n-1}$$

Para una población:

$$\sigma^2 = \frac{\sum_{i=1}^{N} f_i\left(m_i - \mu\right)^2}{N}$$

3.3.2 Desviación típica o estándar

Desviación típica o estándar muestral

$$s = \sqrt{s^2}$$

Desviación típica o estándar poblacional

$$\sigma = \sqrt{\sigma^2}$$

Actividades de Apertura

José, un joven egresado de la universidad, desea realizar un viaje a una ciudad que se encuentre fuera de su país. Ha investigado los costos por noche que implican rentar un cuarto de hotel en tres ciudades diferentes. Al realizar su búsqueda por internet encontró lo siguiente:

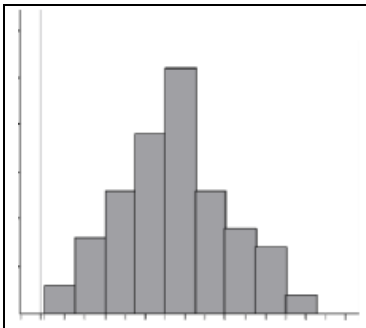
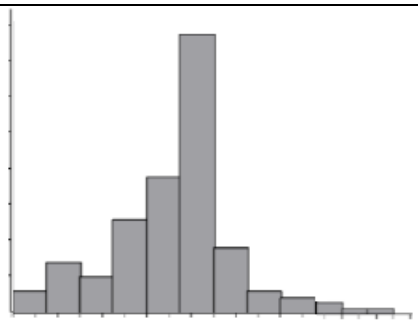
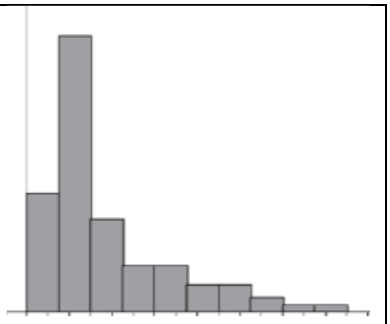
Venecia, Italia. Precio promedio por habitación = 1,865
Nueva York, USA. Precio promedio por habitación = 3,011
Rio de Janeiro, Brasil. Precio promedio por habitación = 797

Al conocer estos datos, el joven se decepcionó al pensar que no podría viajar a Nueva York o Venecia dado el elevado costo del hospedaje por noche; en particular al conocer el costo promedio en cada ciudad, pensó casi de inmediato:

“si el costo promedio por noche es de $3,011 en Nueva York, no tiene caso buscar más opciones de alojamiento en esa ciudad, pues de seguro que todos los cuartos tendrán un costo igual o similar”.

a) ¿Estás de acuerdo con el razonamiento de José?, ¿consideras que debería abandonar su búsqueda por la (supuesta) falta de opciones de alojamientos que sean más económicos?, ¿por qué?

Sintiéndose un poco más optimista, José buscó y revisó más información sobre los costos por noche en el mismo sitio de internet y encontró las siguientes gráficas:

		
$190 – $10,001 MXP Media=1,865 Venecia, Italia	$190 – $15,053 MXP Media=3,011 Nueva York, USA	$190 – $6,063 MXP Media=797 Rio de Janeiro, Brasil

b) ¿Cuál de las tres distribuciones de precios presenta menor variación?

c) ¿Cuál de las tres distribuciones de precios presenta mayor variación?

Actividades de Desarrollo

Supongamos que deseas abrir una cuenta de inversión a tasa variable en una institución financiera. El banco Beta ha ofrecido una tasa de rendimiento del 12, 10, 13, 9 y 11%, mientras que el banco Gama ha proporcionado una tasa del 13, 6, 14, 10 y 12% en los últimos cinco años. Con esta información ¿Cuál banco te ofrece una inversión más segura?

Calcula la varianza de la tasa de rendimiento de los dos bancos y compara los valores. ¿Cambió tu elección?

Calcula la desviación estándar. ¿Qué significa este valor?

Calcular la varianza y desviación estándar de las edades de una población de niños a partir de la siguiente tabla:

Edad (años) x_i	Frecuencia f_i
3	9
4	12
5	9

Actividades de cierre

Pregúntale a 10 de tus compañeros cuánto tiempo demoran en ayudar en las labores del hogar, recomiéndales usar un cronómetro para medir el tiempo.

Calcula la varianza _____ y la desviación estándar_____ del tiempo de recorrido.

Explica el significado de la varianza obtenida _____

_____

Explica el significado de la desviación estándar obtenida.

_____

_____

Actividades de contexto o Transversales

1. Pregúntale a 10 de tus compañeros cuál es su estatura en metros, lleva una cinta métrica para obtener los datos. Analiza y explica los resultados a tus compañeros.

2. Los siguientes son los puntajes de un grupo de adolescentes en un test de Agudeza Visual: 25, 12, 15, 23, 24, 39, 13, 31, 19, 16. Calcular la varianza de estos datos y desviación estándar de los datos.

Ejercicios Adicionales

1. Dada la siguiente distribución de edades en años x, de niños que han padecido varicela.

x	1	2	3	4	5	6	7	8	9	10	11
Frecuencias	4	9	11	8	7	4	5	2	2	0	1

Calcular la desviación estándar.

2. El chef en jefe de cierto restaurante acaba de recibir dos docenas de tomates de su proveedor, pero todavía no los acepta. Sabe por la factura que el peso promedio de un tomate es de 7.5 onzas, pero insiste en que todos tengan un peso uniforme. Aceptará los tomates solo si el peso promedio es de 7.5 onzas y la desviación estándar es menor que 0.5 onzas. Los pesos de los tomates son los siguientes:

6.3	7.2	7.3	8.1	7.8	6.8	7.5	7.8	7.2	7.5	8.1	8.2
8.0	7.4	7.6	7.7	7.4	7.5	8.4	7.4	7.6	7.4	6.2	7.4

¿Cuál es la decisión del chef?

3. Los niños, a diferencia de los adultos, tienden a recordar las películas, cuentos e historias como una sucesión de acciones más que el argumento en forma global y de conjunto. En el relato de una película, por ejemplo, utilizan con frecuencia las palabras "y entonces...". Una psicóloga con suprema paciencia pidió a 25 niños que le contaran una determinada película que ellos habían visto. Consideró la variable: cantidad de "y entonces..." utilizados en el relato y registró los siguientes datos: 15, 19, 17, 18, 17, 16, 16, 21, 23, 20, 20, 15, 17, 20, 22, 17, 19, 20, 18, 11, 31, 17, 19, 18, 40.

Como parte del mismo estudio la experimentadora obtuvo de 25 adultos el mismo tipo de datos. Estos fueron: 10, 5, 13, 10, 8, 7, 9, 7, 9, 10, 3, 11, 14, 8 ,12 ,5 ,10, 9, 7, 11, 14, 10, 15, 9, 4.

Calcular la varianza e indique en cuál grupo los integrantes son más parecidos en cuanto a la cantidad de “y entonces…” utilizados en el relato de una película. Justificar la respuesta.

Varianza del grupo de niños = _____ **Varianza del grupo de adultos =**_____

4. Varianza y desviación estándar para datos agrupados. Completando la siguiente tabla, calcular:

a) La desviación media.

b) La varianza.

c) La desviación típica o estándar.

Intervalos	Frecuencia absoluta f_i	Marca de clase X_i	$X_i \cdot f_i$	Distancia $\lvert X_i - \bar{X} \rvert$	$(X_i - X)^2$
[5.0 - 5.5)	1				
[5.5 - 6.0)	2				
[6.0 - 6.5)	3				
[6.5 - 7.0)	4				
[7.0 - 7.5)	8				
[7.5 - 8.0)	1				
[8.0 - 8.5)	5				
Total					

Bloque 4 | Medidas de forma

4.1 Sesgo

Introducción

Las medidas de forma son indicadores que ofrecen información acerca de la manera en que los datos se encuentran contenidos dentro de una distribución; se clasifican en dos grupos: las medidas de sesgo o asimetría y las medidas de apuntamiento o curtosis.

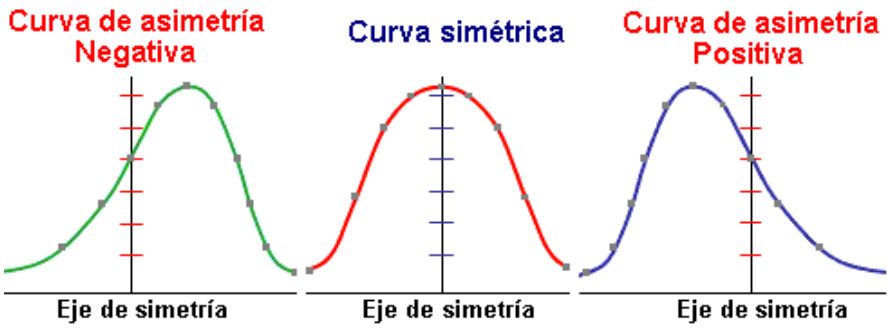
Actividades de Desarrollo

El sesgo es una medida que indica qué tan simétrica o asimétrica es una distribución, se clasifican en:

Distribución simétrica	Se llama así, si los datos que tiene se encuentran repartidos de forma semejante en ambos lados de la media, por lo que en una distribución simétrica la moda, la mediana y la media son iguales.
Distribución asimétrica	Se llama así, si tiene los datos con las frecuencias más bajas ubicadas del lado derecho o izquierdo de la media.

Existen dos tipos de asimetrías

La asimetría a la izquierda	También llamada asimetría negativa, que es aquella en la que los datos con las frecuencias más bajas se ubican a la izquierda de la media y los datos con mayor frecuencia se encuentran a la derecha. En la asimetría negativa la Media es menor que la Mediana y la Mediana es menor que la Moda.
La asimetría a la derecha	También llamada asimetría positiva, y es aquella en la que los datos con las frecuencias más bajas se ubican a la derecha de la media y los datos con mayor frecuencia se encuentran a la izquierda. En la asimetría positiva la Moda es menor que la Mediana y la Mediana es menor que la Media.



Analiza el siguiente vídeo.
https://www.youtube.com/watch?v=88lpvX0YEso

El coeficiente de Fisher es una herramienta que sirve para determinar la simetría o la asimetría de una distribución.

Su fórmula es

$$S_f = \frac{\left(\frac{1}{N}\right)\sum_{i=1}^{k}(X_I - \bar{X})^3}{S^3}$$

N= Número total de datos.
$\bar{X}$= La media aritmética
S = Desviación estándar

El coeficiente de Fisher cumple con las siguientes características.

Si $S_f = 0$ significa que la distribución es simétrica

Si $S_f > 0$ significa que la distribución tiene sesgo positivo

Si $S_f < 0$ significa que la distribución tiene sesgo negativo

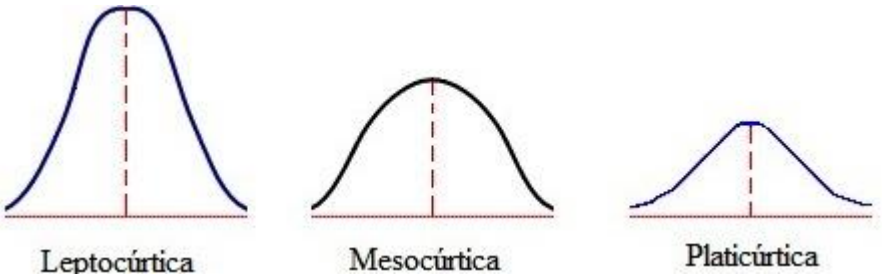
4.2 Apuntamiento o curtosis

El apuntamiento o curtosis mide que tan achatada o tan puntiaguda es una distribución.

De acuerdo con este análisis, las distribuciones se clasifican en tres tipos

Distribución leptocúrtica	Es una distribución que contiene una gran concentración de datos en la zona central.
Distribución mesocúrtica	Es una distribución que contiene una concentración de datos mediana en la zona central.
Distribución platicúrtica	Es una distribución que contiene una baja conglomeración de datos en su región central.

Las siguientes imágenes pueden ayudarte a visualizar los conceptos anteriores.



Actividades de cierre

Contesta las siguientes cuestiones:

1. ¿Qué son las medidas de forma?
2. ¿Qué es sesgo?
3. ¿Qué es una distribución simétrica?
4. Escribe la diferencia que existe entre asimetría positiva y asimetría negativa.
5. Anota ¿qué es el coeficiente de Fisher?
6. De acuerdo con el valor que tiene el coeficiente de Fisher, ¿cómo se clasifican las distribuciones?
7. ¿Qué es apuntamiento o curtosis?
8. ¿Qué es distribución leptocúrtica?
9. ¿Qué es distribución mesocúrtica?
10. ¿Qué es distribución platicúrtica?

Bloque 5 | Medidas de correlación

5.1 Coeficiente de correlación

Introducción

Hasta este momento se han abordado los cálculos de diferentes medidas que nos describen un comportamiento general, su distribución y sus tendencias de una variable. Las medidas de correlación permiten relacionar diversas variables y determinar si tienen o no dependencia alguna entre ellas.

Actividades de Apertura

Revisa la siguiente lección:" Introducción a la Correlación", https://www.youtube.com/watch?v=KNf5soJ4_yk

Actividades de Desarrollo

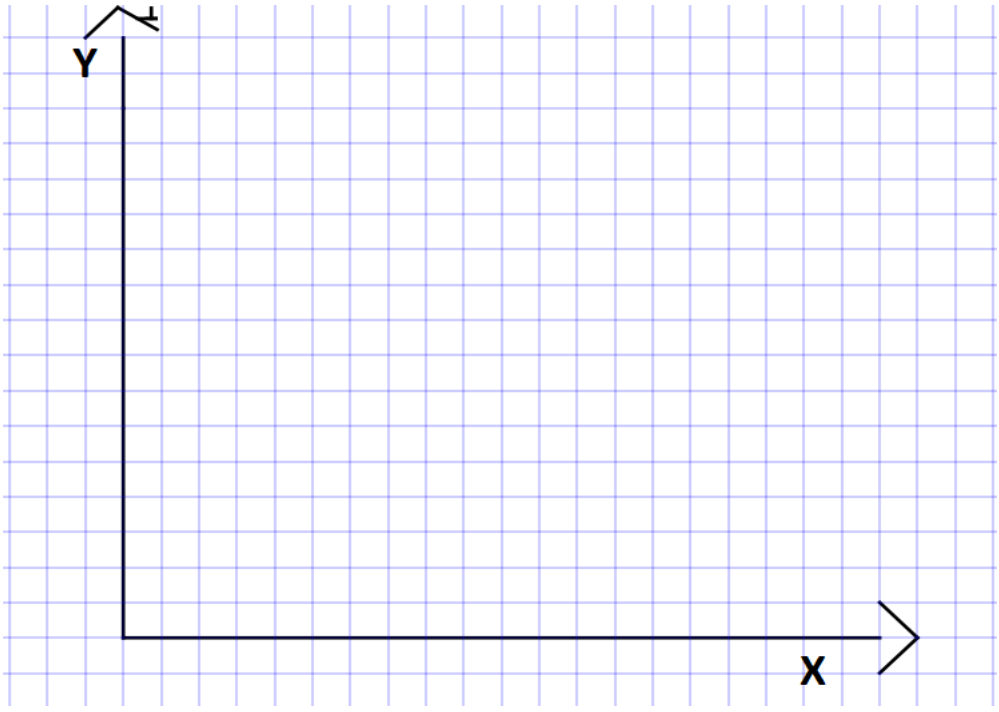
En el departamento de orientación educativa se están organizando conferencias para orientar a los estudiantes en la elección de su carrera a elegir, se buscará la mejor orientación basados en sus calificaciones que determinan sus actitudes y capacidades. Para facilitar la obtención de información al estudiante, se analizarán correlaciones entre sus calificaciones de dos de sus materias, por ejemplo, para el área de físico-matemáticas, se tomarían de Álgebra, Geometría, Cálculo, Física o Química.

Tomemos los siguientes datos de 10 alumnos para ver la relación de sus materias:

Alumno	Cálculo Integral	Física
Alumno 1	48	56
Alumno 2	53	77
Alumno 3	77	82
Alumno 4	44	67
Alumno 5	69	85
Alumno 6	79	86
Alumno 7	84	93
Alumno 8	68	76
Alumno 9	61	58
Alumno 10	55	45

De la tabla de calificaciones, contesta y desarrolla lo que se te pide a continuación:

a) Graficar esta información, usando como abscisas las calificaciones de Cálculo integral y como ordenadas las calificaciones de Física, donde cada punto (x, y), será registrado como la calificación por alumno.



b) Si observas los puntos del gráfico anterior, ¿qué tendencia muestra el conjunto?

Para conocer o dar una respuesta certeza a la pregunta del inciso b), calculemos el coeficiente de correlación “r”:

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

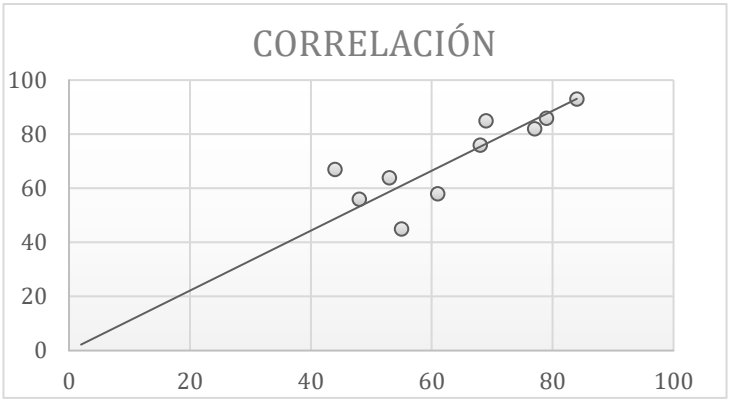
Tomemos Cálculo Integral como la variable X y Física como la variable Y.
Completemos la tabla con la información que necesitamos:

Cálculo X	Física Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
48	56	-15.8	-15.2	240.16	249.64	231.04
53	64	-10.8	-7.2	77.76	116.64	51.84
77	82	13.2	10.8	142.56	174.24	116.64
44	67	-19.8	-4.2	83.16	392.04	17.64
69	85	5.2	13.8	71.76	27.04	190.44
79	86	15.2	14.8	224.96	231.04	219.04
84	93	20.2	21.8	440.36	408.04	475.24
68	76	4.2	4.8	20.16	17.64	23.04
61	58	-2.8	-13.2	36.96	7.84	174.24
55	45	-8.8	-26.2	230.56	77.44	686.44
638	712		**Sumas**	**1568.4**	**1701.6**	**2185.6**
63.8	**71.2**					

$$r = \frac{1568.4}{\sqrt{1701.6} \cdot \sqrt{2185.6}} = \frac{1568.4}{(41.2504545)(46.7504011)} = \mathbf{0.8133}$$

Como podemos observar “r” tiene un valor positivo, quiere decir que las variables tienen una dependencia positiva, es decir, que entre más alta calificación tenga en matemáticas, también subirá la de física. Estos alumnos si son unos buenos candidatos al área de Físico-matemáticas.

Esta es la gráfica correspondiente:



También puedes realizar el ejercicio usando una hoja de Excel, para ello:

Revisa la siguiente lección:" Coeficiente de Correlación",

https://www.youtube.com/watch?v=cRv_Ff315J4&t=15s

Actividades de cierre

Realiza el siguiente ejercicio para ver si existe relación entre las materias de Cálculo Diferencial e inglés.

Alumno	Cálculo Diferencial	Inglés
Alumno 1	44	100
Alumno 2	53	94
Alumno 3	66	90
Alumno 4	69	83
Alumno 5	73	80
Alumno 6	79	78
Alumno 7	81	74
Alumno 8	84	70
Alumno 9	89	68
Alumno 10	95	63

Ejercicios Adicionales

Ejercicio 1: Se midió el peso de 12 personas (x), y la extensión de su cintura (y). Se obtuvieron los siguientes datos.

Persona	1	2	3	4	5	6	7	8	9	10	11	12
x	80	75	90	95	70	65	85	83	78	68	73	92
y	100	90	100	110	90	95	120	120	115	90	99	105

Determina si existe una correlación lineal entre ambas mediciones

	Peso X	Cintura Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
	80	100					
	75	90					
	90	100					
	95	110					
	70	90					
	65	95					
	85	120					
	83	120					
	78	115					
	68	90					
	73	99					
	92	105					

$$r = \frac{\qquad\qquad\qquad\qquad}{\qquad\qquad\qquad\qquad} = \frac{\qquad\qquad\qquad\qquad}{\qquad\qquad\qquad\qquad} =$$

Gráfica

Calcula el coeficiente de correlación o relación que existen entré el peso y la medida de la cintura.

Coeficiente de correlación = _____

Argumentación: _____

_____

Ejercicio 2: Se registraron en la tabla siguiente la edad (en años) y la conducta agresiva (medida en una escala de cero a 10) de 10 niños.

Persona	1	2	3	4	5	6	7	8	9	10
Edad x	6	6	6.7	7	7.4	7.9	8	8.2	8.5	8.9
Conducta y	9	6	7	8	7	4	2	3	2	1

Edad X	Conducta Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
6	9					
6	6					
6.7	7					
7	8					
7.4	7					
7.9	4					
8	2					
8.2	3					
8.5	2					
8.9	1					

Existe relación alguna entre la edad y la conducta agresiva de los niños.

$$r = \frac{\qquad\qquad\qquad\qquad}{\qquad\qquad\qquad\qquad} = \frac{\qquad\qquad\qquad\qquad}{\qquad\qquad\qquad\qquad} =$$

Gráfica

Coeficiente de correlación = _____

Argumentación: _____

_____

5.2 Recta de regresión

Introducción

La recta de regresión intenta sustituir a todo un conjunto de puntos de un diagrama de dispersión y facilitar los cálculos necesarios para determinar la tendencia que presentan las dos variables de interés. No siempre se puede trazar dicha recta de regresión para determinar la tendencia. Esta recta también nos sirve para extrapolar y conocer datos de que no se encuentran en la tabla de información.

Actividades de Apertura

Revisa la siguiente lección: “Recta de Regresión Lineal”,
https://www.youtube.com/watch?v=9U-cCKW-gDA&t=12s

Actividades de Desarrollo

Es hora de calcular la recta de regresión lineal de un conjunto de datos.

Sea la ecuación de la recta:

$$y = bx + a + €$$

Donde:

b= Es la pendiente de la recta, este valor debe estimarse a partir de la tendencia de los puntos

a= Es la ordenada al origen, es decir, el punto en que la recta cruza el eje Y

y= Es la variable dependiente

x= Es la variable independiente

€ es el error

$\boldsymbol{b = \frac{COVxy}{Sx}}$ donde S_x es la Varianza de x

$\boldsymbol{a = \bar{y} - b\bar{x}}$ donde $\bar{y}$ es la media de la variable y, $\bar{x}$ es la media de la variable **x**

Alumno	1	2	3	4	5	6	7	8	9	10
Cálculo (x)	48	53	77	44	69	79	84	68	61	55
Física (y)	56	64	82	67	85	86	93	76	58	45

$$COVxy = \frac{\sum_1^n (x_i - \bar{x})\,(y_i - \bar{y})}{\mathrm{n} - 1} \qquad Sx = \frac{\sum_1^n (x_i - \bar{x})^2}{\mathrm{n} - 1}$$

Ahora tomemos los datos de la tabla

Cálculo X	Física Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
48	56	-15.8	-15.2	240.16	249.64	231.04
53	64	-10.8	-7.2	77.76	116.64	51.84
77	82	13.2	10.8	142.56	174.24	116.64
44	67	-19.8	-4.2	83.16	392.04	17.64
69	85	5.2	13.8	71.76	27.04	190.44
79	86	15.2	14.8	224.96	231.04	219.04
84	93	20.2	21.8	440.36	408.04	475.24
68	76	4.2	4.8	20.16	17.64	23.04
61	58	-2.8	-13.2	36.96	7.84	174.24
55	45	-8.8	-26.2	230.56	77.44	686.44
638	712		**Sumas**	**1568.4**	**1701.6**	**2185.6**
63.8	**71.2**					

Vamos a sustituir los valores:

$$COV_{XY} = \frac{1568.4}{9} = 174.266667$$

$$S_X = \frac{1701.6}{9} = 189.066667$$

$$b = \frac{174.267}{189.067} = 092172073$$

$$a = 71.2 - (0.9217207)(63.8) = 12.394217$$

Entonces obtenemos la ecuación de la recta:

$$\boldsymbol{y = 0.9217x + 12.3942 + €}$$

Actividades de cierre

Encuentra la ecuación de la Recta Lineal del ejercicio anterior.

Alumno	Cálculo Diferencial	Inglés
Alumno 1	44	100
Alumno 2	53	94
Alumno 3	66	90
Alumno 4	69	83
Alumno 5	73	80
Alumno 6	79	78
Alumno 7	81	74
Alumno 8	84	70
Alumno 9	89	68
Alumno 10	95	63

Ejercicios Adicionales

1. Rescata el ejercicio de la actividad de cierre del tema de Coeficiente de correlación referente a los pesos y medidas de cintura para calcular la ecuación de la recta de regresión correspondiente.

Peso X	Cintura Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
80	100					
75	90					
90	100					
95	110					
70	90					
65	95					
85	120					
83	120					
78	115					
68	90					
73	99					
92	105					

$COV_{XY} =$

$S_X =$

$b =$

$a =$

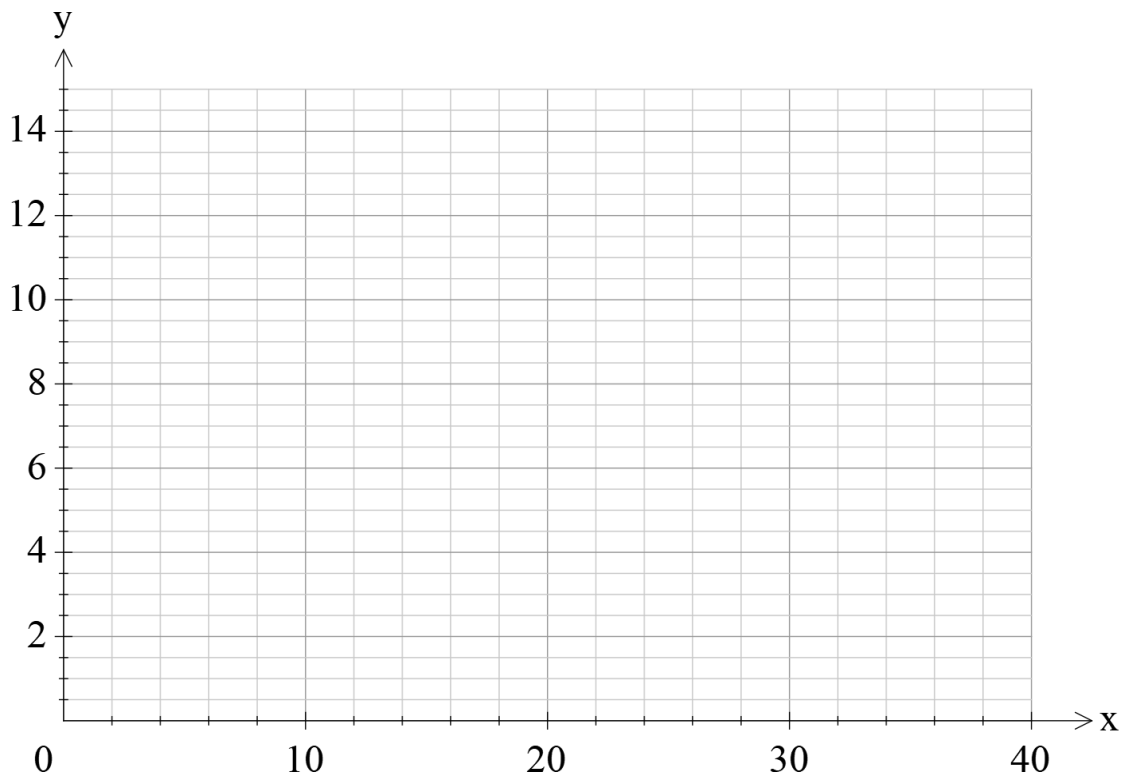
Por tanto, la recta queda de la siguiente forma:

$\boldsymbol{y} =$

2. Un centro comercial sabe que, en función de la distancia, en kilómetros, a la que se sitúe de un núcleo de población, acuden los clientes, en cientos (por día), que figuran en la tabla:

Caso	N° de Clientes (en cientos) y	Distancia (en Km) x	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	8	15					
2	7	19					
3	6	25					
4	4	3					
5	2	34					
6	1	40					
Sumas							
Medidas							

a) Construye el grafico de dispersión de las variables y grafica la recta de regresión



Cálculo de los Coeficientes de Regresión:

a=

b=

Ecuación de la Recta de regresión:

Estima el número de clientes que debe esperarse si el Centro Comercial se construye a 2 km de la población:

Calcula el Coeficiente de Correlación:

¿Qué interpretación se da al valor coeficiente de correlación?:

5.3 Error estándar de estimación

Introducción

El error de estimación nos mide el grado de alejamiento de existe entre cada punto del diagrama de dispersión y la recta de regresión. Debido a que la recta no pasa por cada uno, si así sucediera no sería entonces una recta, se dice entonces que se comete un error al definir la Recta de Regresión Lineal.

Actividades de Apertura

Revisa la siguiente lección: “Error Estándar”, https://www.youtube.com/watch?v=LWbVyDzmIaA&t=11s

Actividades de Desarrollo

Siguiendo con el ejercicio inicial, definamos el error mediante la siguiente ecuación:

Recordemos que:

$$€ = \sqrt{\frac{\sum_1^n (y_i - \bar{y})^2}{N-2}}$$

Lo tomamos de la primera tabla donde hicimos los cálculos

$$€ = \sqrt{\frac{2185.6}{10-2}} = 16.53$$

El error se obtiene del cálculo de una raíz por tanto su valor es **“+” “-”**

Así nos quedaría la ecuación de la Recta de regresión lineal:

y = 0.9217x + 12.3942 ±16.529

Actividades de cierre

Calcula la estimación del error del ejercicio de la recta de regresión lineal que encontraste en la Actividades de **cierre del punto 5.2.**

Ejercicios Adicionales

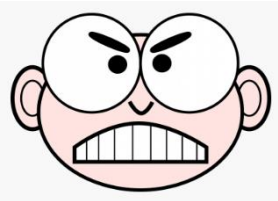
Ejercicio 1: Se midió el peso de 12 personas (x), y la extensión de su cintura (y). Se obtuvieron los siguientes datos.

Persona	1	2	3	4	5	6	7	8	9	10	11	12
x	80	75	90	95	70	65	85	83	78	68	73	92
y	100	90	100	110	90	95	120	120	115	90	99	105

Calcula el error de la estimación de la recta de Regresión Lineal que se aproxima a este conjunto de datos.

Error = _____

Ejercicio 2: Se registraron en la tabla siguiente la edad (en años) y la conducta agresiva (medida en una escala de cero a 10) de 10 niños.



Persona	1	2	3	4	5	6	7	8	9	10
Edad x	6	6	6.7	7	7.4	7.9	8	8.2	8.5	8.9
Conducta y	9	6	7	8	7	4	2	3	2	1

Calcula el error de la estimación de la recta de Regresión Lineal que se aproxima a la relación entre la edad y la conducta de los niños.

Error = _____

Ejercicio 3: Calcula el error estándar de estimación del ejercicio 2 de la página 113.

Bloque 6 | Teoría de conjuntos

6.1 Elementos básicos

6.1.1 Diagrama de Venn-Euler

Introducción

Te habrás dado cuenta de que los conjuntos también se pueden representar gráficamente mediante diagramas de Venn- Euler, que consisten en círculos que contienen en un interior los elementos del conjunto. En el estudio de conjuntos es necesario establecer uno de los conjuntos del cual proceden todos los conjuntos con los cuales se efectúan las operaciones. Por ejemplo, la mayoría de los conjuntos numéricos puede usarse como origen el conjunto de los números reales. Al conjunto que se considera como el origen de todos los conjuntos con los que estamos trabajando se le denomina conjunto universo y lo podemos representar con la letra "U" mayúscula, como ya se mencionó anteriormente. Este conjunto Universal será un rectángulo, que encierre todos los conjuntos representados por círculos, así podemos comprender que el complemento de un conjunto contiene todos los elementos de ese conjunto y que también pueden estar en el universo.

Actividades de Apertura

Analiza en tu casa las cosas que te rodean, podrás observar que están acomodadas de tal manera que sea fácil encontrarlas. Por ejemplo, en el refrigerador, encontrarás la fruta en un lado, en otro espacio encontrarás la leche o jugos, en otro espacio, la carne, etc.
De igual manera los diagramas de Venn- Euler nos sirven para acomodar los conjuntos y establecer relaciones que tienen entre ellos.

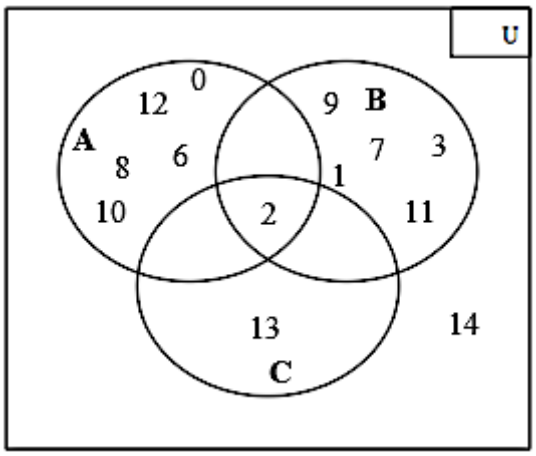
Dibuja en tu cuaderno el refrigerador de tu casa y como ordenarías lo que hay adentro.

Actividades de Desarrollo

Dado los siguientes conjuntos:

$$U = \{0,1,2,3,4,5,6,7,8,9,10,11,12,13,14\}$$
$$A = \{0,2,4,6,8,10,12\}$$
$$B = \{1,2,3,5,7,9,11\}$$
$$C = \{2,4,5,13\}$$

El diagrama de Venn- Euler del sistema formado con estos conjuntos es:



Actividades de cierre

1. Piensa en un diagrama de Venn- Euler que puedas realizar con los siguientes conjuntos y desarróllalo en tu libreta.

$$U = \{-1, -2, -3, -4, -5, -6\}$$
$$A = \{-1, -2, -3\}$$
$$B = \{-1, -4, -6\}$$
$$C = \{-1, -3\}$$

6.1.2 Conjunto

Introducción

Nuestro entorno está formado por conjuntos, por ejemplo; tu familia, escuela, casa, ropa, las estrellas, planetas, jugadores favoritos de algún deporte de nuestro gusto, etc., todos ellos son conjuntos. Dado que los conjuntos aparecen en todo nuestro entorno, es de vital importancia saber expresarlos y conocer los procedimientos para relacionar sus elementos con los de otros conjuntos.

En Matemáticas, los números y las cantidades son conjuntos; en Geometría o en Trigonometría las figuras geométricas también son conjuntos y tienen elementos en común, por ello es necesario conocer reglas, formas y operaciones que permitan trabajar con ellos y que sea posible usarlos en las aplicaciones cotidianas del ser humano.

Un conjunto es una colección de elementos organizados por alguna característica en común.

Es importante que el conjunto este definido sin ambigüedades, respecto a los elementos que forman parte de él.
Conjunto: ”Clase de objetos definida apropiadamente”ii

Los conjuntos se definen con letras mayúsculas, a las cuales se asigna una expresión entre llaves y sus elementos se separan con comas. Como este es un tema muy amplio, tu facilitador te explicará las propiedades de los conjuntos y sus operaciones con más detalle.

Actividades de Apertura

Observa dentro de tu mochila, las cosas que guardas en ella se organizan de acuerdo a ciertas características, por ejemplo, en tu lapicera guardas, lápices, lapiceros, colores, sacapuntas, borrador y cosas que te servirán en clases, en otro espacio de tu mochila guardas tus libretas de las diversas materias, en otro espacio tus libros, etc.
Como verás existen conjuntos de cosas que guaramos ordenadamente considerando ciertas características en común.
Ahora observa el video y analízalo, esto permitirá entender el concepto básico sobre conjuntos.

Observa el video y analiza.
https://www.youtube.com/watch?v=PZN6jyw1Nqk

Actividades de Desarrollo

Recordemos que los conjuntos se pueden definir por Comprensión o por Extensión.
Por extensión cuando enumeras cada uno de los elementos.
Ejemplo.

$$A = \{1,2,3,4,5,6\}$$

Por comprensión cuando describes el conjunto por alguna característica en particular o en general.

Ejemplo:

$$A = \{x/x \in N, x < 7\}$$

Conjunto bien definido.
V es el conjunto de vocales $V = \{a, e, i, o, u\}$
Conjunto que no está bien definido.

$M = \{10\ mejores\ deportistas\}$
Este segundo ejemplo no está bien definido por qué no se puede determinar ¿En qué deporte o deportes? o ¿Quiénes son?

Rosita tiene un conjunto de frutas ella desea clasificar la fruta para colocarla dentro de su refrigerador, pretende colocarla según su estructura.

Observa el conjunto U=Universo de fruta que tiene Rosita para clasificar su fruta de la siguiente manera:

$U = \{Melón, Sandía, mango, uvas, fresas, plátono, manzanas, peras,\}$

Escribe en el conjunto A, las frutas de cáscara suave y en el conjunto B las frutas de cáscara dura.
Si no tienes el espacio suficiente escribe esta actividad de desarrollo en tu libreta de apuntes y en la libreta de ejercicios inventa otros ejemplos como este.

$A = \{$ $\}$
$B = \{$ $\}$

Si Rosita quisiera agregar un conjunto donde indique frutas sin cáscara, cuales frutas escribiría dentro del conjunto C.

$C = \{$ $\}$

Actividades de cierre

Pon a prueba tu ingenio y escribe por lo menos 10 conjuntos que observes a tu alrededor. Realízalo en tu libreta de apuntes.

Resuelve el siguiente ejercicio y escribe los siguientes conjuntos por extensión en tu libreta.

Conjunto P de planetas del sistema solar.
$P = \{$ $\}$

Conjunto de números Naturales.
$N = \{$ $\}$

Conjunto de números Enteros.
$Z = \{$ $\}$

Conjunto de números pares menores que 10.
$A = \{$ $\}$

Conjunto de números impares mayores que -7 y menores que 9.
$B = \{$ $\}$

De que otra manera podría escribir este último ejemplo por comprensión.
$B = \{$ $\}$

6.1.3 Subconjunto

Introducción

Cuando todos los elementos de un conjunto A son elementos contenidos en otro conjunto B, se dice que A es subconjunto de B. Esto se denota mediante la expresión $A \subset B$

Actividades de Apertura

Un subconjunto es una parte de un conjunto o bien que el subconjunto está incluido dentro de un conjunto más grande.

Si tu observas las cosas que tienen en la mochila, podrás observar que tienes varios subconjuntos dentro de la mochila, como ya habíamos analizado anteriormente. En una de las bolsas de tu mochila por ejemplo tienes otra bolsa donde tienes un subconjunto ejemplo monedero en el cual están incluidos todos los elementos de ese conjunto y ahí encontrarás tal vez monedas y billetes.
Observa el video y analiza.

YouTube https://www.youtube.com/watch?v=e31Th_hFB5c

Actividades de Desarrollo

Analicemos un ejemplo sobre subconjuntos.
Sean los conjuntos A y B

$$A = \{1,3,5,7,18\}$$
$$B = \{1,7,18\}$$

Donde podemos observar que:

$$B \subseteq A$$

B es subconjunto de A

La línea debajo del símbolo ⊆, significa que B puede ser igual que A, estos pueden ser conjuntos idénticos. Si queremos decir que B es un subconjunto apropiado de A, esto quiere decir: es un subconjunto, pero hay por lo menos un elemento en A que no está en B, entonces debemos escribirlo correctamente eliminando la línea de abajo y quedará de la siguiente manera:

$$B \subset A$$

Entonces para otro ejemplo sería:

$$M = \{a, e, i, o, u\}$$
$$N = \{a, e, i, o, u\}$$

Y nos piden determinar que N es subconjunto de M, puedo expresarlo así:

$$N \subseteq M$$

Porque estaré indicando que N es un subconjunto de M pero que también es igual a M.

Actividades de cierre

Analiza los siguientes conjuntos indica quien es el subconjunto.

Sean los conjuntos F y G

$$\boldsymbol{F = \{Figuras\ geométricas\}}$$
$$\boldsymbol{G = \{Triángulos, cuadriláteros\}}$$

Sean los conjuntos J y F

$J = \{Triángulos\}$

$F = \{triángulos\ rectangulos, triangulos\ isosceles\}$

Escribe como representarías que F es subconjunto de J

Y represéntalo en un diagrama de Venn Euler.

U

Puedes desarrollar 5 ejercicios sobre este tema, puedes buscarlos en libros de tu biblioteca o bien virtuales que puedas encontrar.

6.1.4 Pertenencia

Introducción

Un elemento pertenece a un conjunto, cuando éste, está contenido dentro del conjunto de referencia.
El símbolo que se usa para indicar pertenencia es:

$$\in$$

Por el contrario, cuando un elemento no pertenece a un conjunto, quiere decir que ese elemento no está contenido en el conjunto de referencia.

Símbolo de no pertenece es:

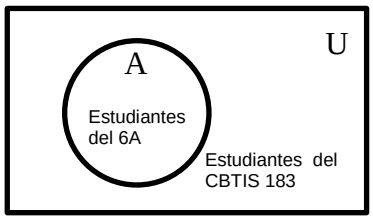
$$\notin$$

Actividades de Apertura

Supongamos que tu escuela la observamos como un conjunto Universo, el grupo en el que estás inscrito(a) pertenece a tu escuela, esto nos indica que tu grupo está incluido dentro de ese grupo Universo.

Ejemplo si tú eres un estudiante del 6° A del CBTIS 183, podríamos decir que el 6°A pertenece al CBTIS 183.

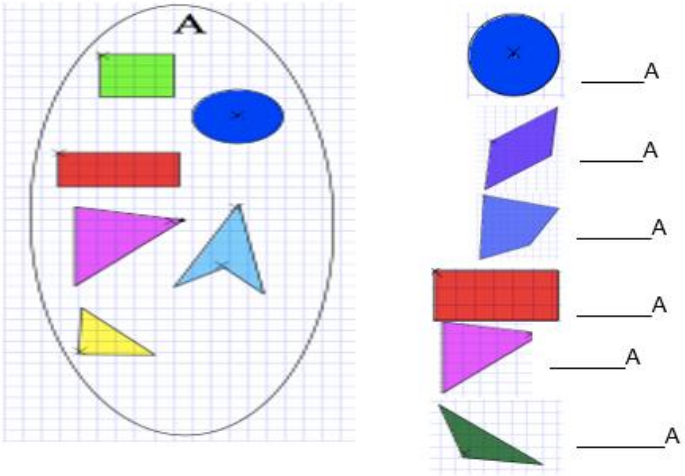
$$U = \{estudiantes\ del\ CBTIS\ 183\}$$
$$A = \{estudiantes\ del\ 6A\ de\ la\ especialidad\ de\ electricidad\}$$



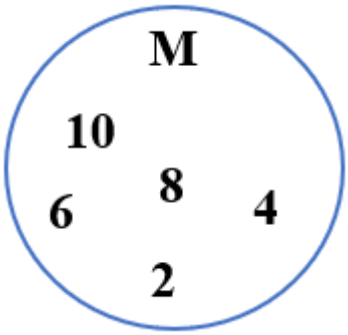
Observa el video y analiza.
https://www.youtube.com/watch?v=C260x7UApNg

Actividades de Desarrollo

Observa el siguiente conjunto A, e identifica que figura pertenece al conjunto y escribe el símbolo de pertenece o no pertenece según sea el caso.



Este es otro ejemplo con números, escribe los símbolos de “pertenece” o “no pertenece”.



2 _____ M

3 _____ M

4 _____ M

12 _____ M

Actividades de cierre

Dibuja en tu libreta un ejemplo parecido, puedes utilizar otros conjuntos de figuras, objetos, cosas, personas, etc.

6.1.5 Conjunto universo

Introducción

El conjunto universo, es aquel que contiene a todos los elementos de interés, mayormente se representa con la letra U mayúscula.

Actividades de Apertura

Imagina a todos los planteles que pertenecen a la UEMSTIS, podemos considerarlo como un conjunto universo, es finito porque tiene un número limitado de planteles, lo podemos considerar como conjunto Universal.

Si hablamos de personas en el planeta tierra podemos considerarlo como conjunto Universal o Conjunto Universo de igual manera porque existen millones de personas en el planeta Tierra.

Existen Conjuntos Universos que están limitados y otros que son infinitos, como el conjunto Universal de las estrellas en el Universo estelar. Entonces podemos observar que el Conjunto Universo puede estar determinado por un límite de elementos o bien de manera ilimitada.

Observa el video y analiza.
https://www.youtube.com/watch?v=L_J0UOf_u5s

En el siguiente ejemplo se presenta el conjunto universo

$$\boldsymbol{U = \{x/x\ \textit{es una letra de nuestro alfabeto en México}\}}$$

Representa de otra forma el universo anterior, recuerda que puede ser por comprensión y por extensión, desarróllalo en tu libreta.

$$\boldsymbol{U = \{ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \}}$$

Actividades de cierre

Piensa y escribe en tu libreta, en 5 diferentes universos no muy extensos en los que puedas representarlos por extensión y por comprensión.

6.1.6 Conjunto vacío

Introducción

El conjunto vacío, es el conjunto que no contiene elementos. Se representa con las siguientes expresiones o símbolos:

$$\emptyset \quad \{\,\}$$

Actividades de Apertura

Analicemos al conjunto vacío. En esta contingencia, analicemos el mes de mayo de 2020, si te pregunto ¿En este momento cuántos estudiantes están en clases normales? La respuesta será ningún estudiante porque todos están en sus casas realizando sus tareas y trabajos. Podemos afirmar que si le asignamos una letra para determinar el conjunto podemos decir:

$$A = \{Estudiantes\ que\ están\ en\ clase\ en\ el\ mes\ de\ mayo\ 2020\}$$

Este conjunto es vacío $\boldsymbol{A = \{\ \}}$

Observa el video y analiza.
https://www.youtube.com/watch?v=L_J0UOf_u5s

Actividades de Desarrollo

Un ejemplo del conjunto Vacío se presenta cuando queremos definir al conjunto de todos los números naturales que sean negativos, es decir: $\boldsymbol{B = \{x/x < 0, x \in N\}}$

Sabemos que los números Naturales se definen como enteros positivos, por lo que en este conjunto no pueden existir negativos, en consecuencia $B = \emptyset$

Actividades de cierre

En un experimento lanzamos un dado, representamos los posibles resultados con los elementos del conjunto R.

$$R = \{1,2,3,4,5,6\}$$

$$F = \{al\ lanzar\ el\ dado\ obtengo\ el\ número\ 9\}$$

¿Cómo es el conjunto resultante?:

Escribe tu propia conclusión sobre el conjunto vacío:

6.2 Operaciones con conjuntos

6.2.1 Unión

Introducción

La unión de conjuntos, consiste en formar un conjunto que contenga todos los elementos de los conjuntos que se unen. Así la unión del conjunto A con el conjunto B es el conjunto que tiene todos los elementos del conjunto A y todos los elementos de B. El símbolo de la unión es ∪ de tal manera que, para representar la unión de los conjuntos A y el conjunto B, se representa $A \cup B = B \cup A$

Actividades de Apertura

Cuando se unen dos conjuntos, el nuevo conjunto unión, está formado por los elementos de ambos conjuntos, quiere decir que pertenecen al primer conjunto o al segundo conjunto.
Si considero dos conjuntos A y B como conjuntos se puede definir de la siguiente manera:

$$\boldsymbol{A \cup B = \{x / x \in A \vee x \in B\}}$$

Observa el video y analiza.
https://www.youtube.com/watch?v=MSxkIyHOCvA

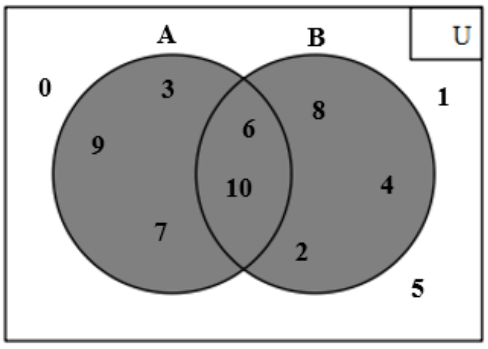
Actividades de Desarrollo

Sean los siguientes conjuntos:

$U = \{0,1,2,3,4,5,6,7,8,9,10\}$
$A = \{3,6,7,9,10\}$
$B = \{2,4,6,8,10\}$

Entonces

$$A \cup B = \{2,3,4,6,7,8,9,10\}$$



Esta unión se puede representar mediante un diagrama de Venn- Euler.

$$A \cup B = \{2,3,4,6,7,8,9,10\}$$

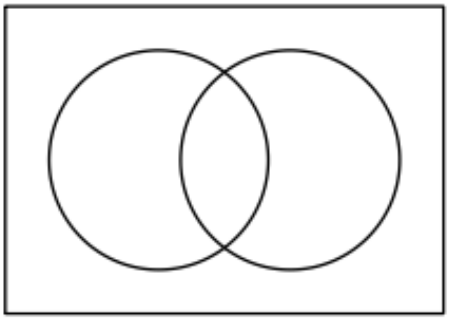
Actividades de cierre

Desarrolla en tu libreta el siguiente ejercicio, tu facilitador te proporcionará otros ejercicios para que puedas practicar.

1.- Dados los conjuntos:

U={1,2,3,4,5,6,7,8,9,10,11,12}
A={2,4,6,8,10,12}
B={1,2,3,5,7,11}

$A \cup B = \{$ $\}$
Represéntalo en el diagrama de Venn- Euler

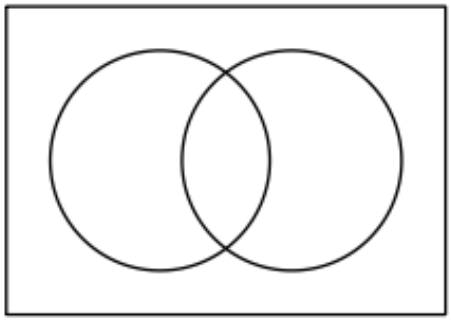


2.-Dado los siguientes conjuntos

$U = \{letras\ del\ alfabeto\}$
$M = \{a, e, i, o, u\}$
$N = \{a, m, e, r, i, c, a\}$

$M \cup N = \{$ $\}$

Represéntalo en el diagrama de Venn- Euler



6.2.2 Intersección

Introducción

La intersección de conjuntos, es el conjunto formado con todos los elementos de un conjunto que también pertenecen a otro conjunto, es decir, la intersección del conjunto A con el conjunto B es el conjunto de todos los elementos del conjunto A que pertenecen al conjunto B y viceversa. El símbolo de la intersección es ∩

La intersección se expresa de la siguiente manera $\mathbf{A} \cap \mathbf{B}$. También se cumple que $\mathbf{A} \cap \mathbf{B} = \mathbf{B} \cap \mathbf{A}$

Actividades de Apertura

Cuando tenemos dos conjuntos, ejemplo el conjunto A y el Conjunto B, la intersección de estos dos conjuntos, es un nuevo conjunto en el cual están contenidos los elementos comunes de ambos conjuntos, lo cual quiere decir que están tanto en el Conjunto A y en el conjunto B.

Se puede definir de la siguiente manera:

$$\boldsymbol{A} \cap \boldsymbol{B} = \{\boldsymbol{x}/\boldsymbol{x} \in \boldsymbol{A} \wedge \boldsymbol{x} \in \boldsymbol{B}\}$$

Observa el video y analiza.
https://www.youtube.com/watch?v=11Z2gOSyl7k&t=6s

Actividades de Desarrollo

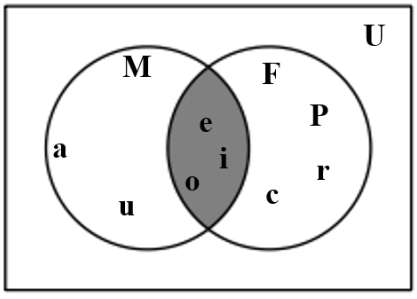
Sea los conjuntos:

$$M = \{a, e, i, o, u\}$$
$$F = \{p, e, r, i, c, o\}$$

La intersección de los dos conjuntos es:

$$M \cap F = \{e, i, o\}$$

Se puede representar mediante un diagrama de Venn- Euler.

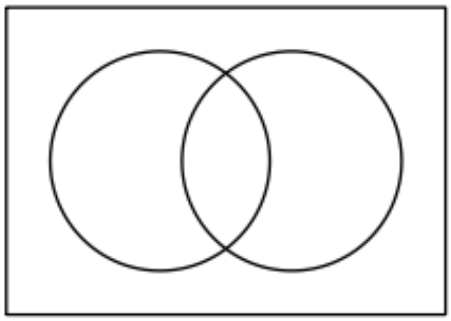


$$M \cap F = \{e, i, o\}$$

Actividades de cierre

Analiza el siguiente ejercicio y determina el conjunto intersección, puedes trabajar en tu libreta.

1. Sean los conjuntos Universo U y conjuntos A y B.

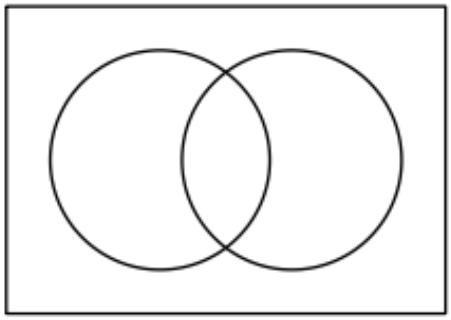


$U = \{1,2,3,4,5,6,7,8,9,10\}$
$A = \{x/x\ es\ un\ número\ par, x \in U\}$
$B = \{x/x\ es\ un\ número\ impar, x \in U\}$

Determina: $\boldsymbol{A} \cap \boldsymbol{B} =$

Y representa su diagrama de Venn- Euler.

2.- Sean los conjuntos Universo U y conjuntos F y G



$U = \{a, b, c, d, e, f, g, h, i, j, k, l, ll, m, n, ñ, o, p, q, r\}$
$F = \{a, m, o, r\}$
$G = \{p, e, r, o\}$

Determina: $\boldsymbol{F} \cap \boldsymbol{G} = \{$ $\}$

Y representa su diagrama de Venn- Euler.

6.2.3 Complemento

Introducción

Sea A un subconjunto, de un conjunto universal U. El complemento del conjunto A, considerando el Universo U, es el conjunto A' o A^c, de todos los elementos del universo U, que no pertenecen al conjunto A.

Actividades de Apertura

El conjunto complemento se refiere a comprender que si en un conjunto ejemplo A, deseamos determinar su complemento, este nuevo conjunto complemento son los elemento que están en el Universo y no pertenecen al conjunto A, se puede definir como se expresa a continuación:

$$A^c = A' = \{x / x \in U \wedge x \notin A\}$$

Observa el video y analiza.
https://www.youtube.com/watch?v=yjRO-1n6OUo

Sea el conjunto Universo y el conjunto A:

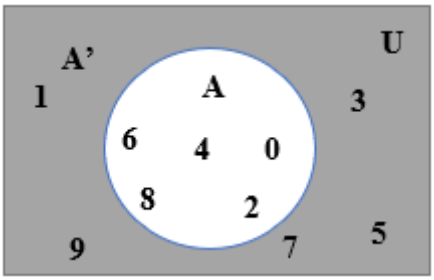
$$U = \{0,1,2,3,4,5,6,7,8,9\}$$
$$A = \{0,2,4,6,8\}$$

Su complemento es:

$$A^c o\ A' = \{1,3,5,7,9\}$$

Esto se puede representar en un diagrama de Venn Euler.

$$A^c o\ A' = \{1,3,5,7,9\}$$



Actividades de cierre

Analiza el siguiente ejercicio y determina el conjunto complemento, puedes trabajar en tu libreta.

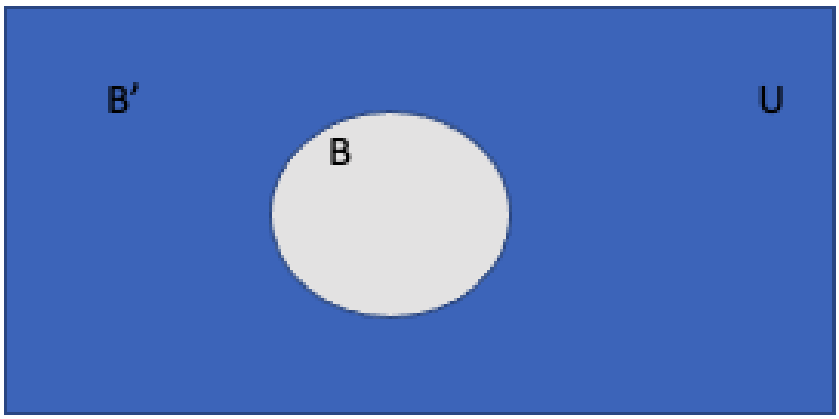
1.- Sean los conjuntos Universo U y conjuntos A y B.

$$U = \{x/0 \leq x \leq 10\,, x \in N\}$$

$$B = \{2,4,6,8\}$$

Determina el conjunto complemento B' y represéntalo mediante un diagrama de Venn- Euler.

$$B' = \{\qquad\qquad\qquad\qquad\}$$

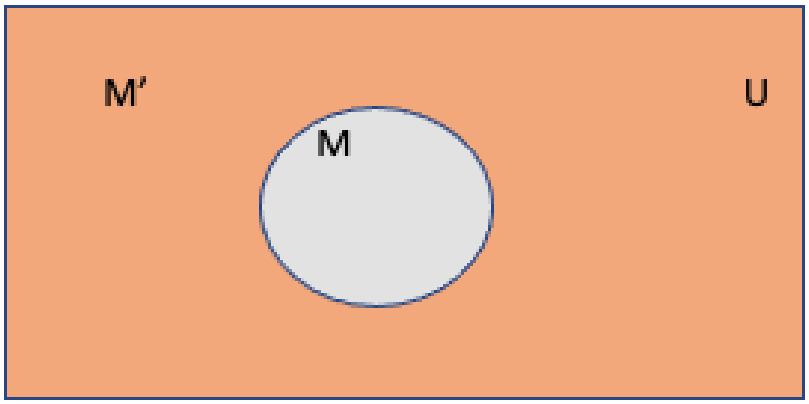


2. Sean los conjuntos Universo y conjunto M.

$U = \{x/2 < x \leq 8, x \in N\}$
$M = \{3,4,5,6,7,8\}$

Determina el conjunto complemento M' y represéntalo mediante un diagrama de Venn- Euler.

$$M' = \{\qquad\qquad\qquad\qquad\}$$



6.2.4 Diferencia

Introducción

La diferencia entre los conjuntos A y B, se pueden representar de la siguiente manera:
A-B o B-A, que dan lugar a conjuntos diferentes.
A-B es el conjunto de todos los elementos de A que no pertenecen a B, mientras que B-A es el conjunto de todos los elementos de B que no pertenecen al conjunto A.

Actividades de Apertura

Si consideramos al conjunto A y al conjunto B, la diferencia A-B, es el conjunto de elementos de A que no son elementos de B determinándolos de la siguiente manera:

$$\boldsymbol{A - B = \{x / x \in A \wedge x \notin B\}}$$

Por el contrario B-A se puede entender de la siguiente manera:

$$\boldsymbol{B - A = \{x / x \in B \wedge x \notin A\}}$$

YouTube

Observa el video y analiza.
https://www.youtube.com/watch?v=GGLmJityzcw

Actividades de Desarrollo

Sean los conjuntos

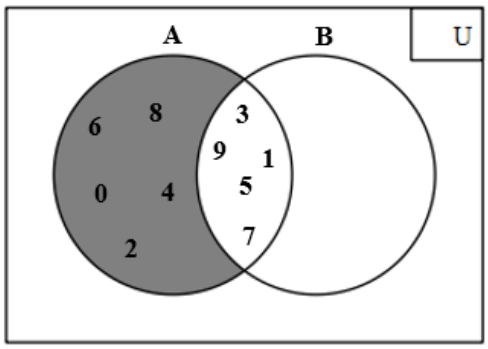
$$A = \{0,1,2,3,4,5,6,7,8,9\}$$

$$B = \{1,3,5,7,9\}$$

$$A - B = \{0,2,4,6,8\}$$

Esto se puede representar con un diagrama de Venn-Euler.

$$A - B = \{0,2,4,6,8\}$$



Proponemos otro ejemplo:

Sea el siguiente Universo U, conjuntos M y N.

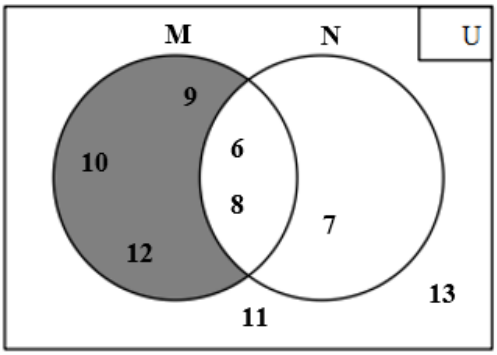
$$U = \{6,7,8,9,10,11,12,13\}$$

$$M = \{6,8,9,10,12\}$$

$$N = \{6,7,8\}$$

$$M - N = \{9,10,12\}$$

$$M - N = \{9,10,12\}$$

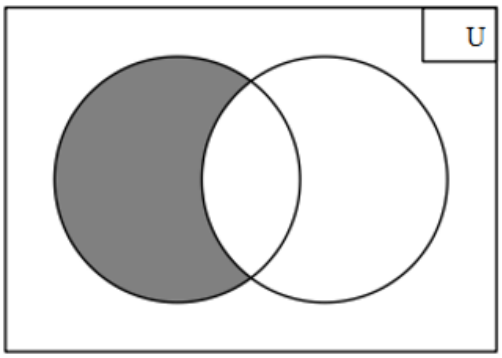


Actividades de cierre

Analiza y desarrolla en tu libreta el siguiente ejercicio.

1. Dados los siguientes conjuntos.

$U = \{letras\ de\ nuestro\ \ alfabeto\}$
$M = \{a, e, i, o, u\}$
$N = \{a, m, e, r, i, c, a\}$
$M - N =$



6.3 Representación de operaciones mixtas con hasta tres conjuntos

Introducción

Para el desarrollo de las operaciones mixtas se requiere que el estudiante domine los temas analizados anteriormente, comprendiendo los elementos y operaciones con conjuntos, puesto que esos elementos básicos permitirán encontrar las soluciones de este tipo de ejercicios.

Actividades de Apertura

En este último tema analizado, nos permitirá observar como podemos trabajar con tres conjuntos diferentes y que tienen algunas similitudes o números, objetos, elementos similares, lo cual nos permitirá poner en práctica todo lo analizado en los temas anteriores.

Aquí podrás poner en juego tu capacidad cognitiva y de orden superior para poder analizar ejercicios que contienen tres conjuntos y donde aparecen varias operaciones de las analizadas anteriormente.

Tu facilitador podrá explicarte con más detalle el tema y podrás realizar otros ejercicios que te permitan desarrollar tu creatividad y capacidad para analizar y proponer ejemplos de este tipo.

Actividades de Desarrollo

Sean los siguientes conjuntos:

$$U = \{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15\}$$

$$A = \{x/x = 2n - 1, n \in N, x \in U\}$$

$$B = \{x/x\ es\ un\ número\ primo\ , x \in U\}$$

$$C = \{4,5,6,8,9,11,12\}$$

Realicemos la siguiente operación.

$$\boldsymbol{B' \cap (C \cup A)' =}$$

Este ejercicio lo podemos desarrollar de la siguiente manera:

Retomamos los conjuntos de una manera que se pueda entender mejor.

$$U = \{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15\}$$

$$A = \{1,3,5,7,9,11,13,15\}$$
$$B = \{2,3,5,7,11,13\}$$
$$C = \{4,5,6,8,9,11,12\}$$

$B' \cap (C \cup A)' =$

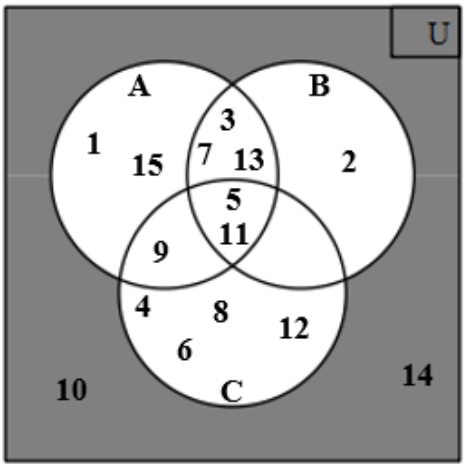
$B \cap (C \cup A) = \{2,3,5,7,11,13\} \cap \{4,5,6,8,9,11,12\} \cup \{1,3,5,7,9,11,13,15\}$

$B' \cap (C \cup A) = \{1,4,6,8,9,10,12,14,15\} \cap \{1,3,4,5,6,7,8,9,11,12,13,15\}$

$\boldsymbol{B' \cap (C \cup A)' = \{1,4,6,8,9,10,12,14,15\} \cap \{2,10,14\} = \{10,14\}}$

Diagrama de Venn – Euler.

$$\boldsymbol{B' \cap (C \cup A)' = \{10, 14\}}$$



Actividades de cierre

Con los mismos conjuntos de la explicación anterior, determina las siguientes operaciones y represéntalas en un diagrama de Venn – Euler.

Retomemos nuevamente los conjuntos:

$$U = \{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15\}$$

$$A = \{1,3,5,7,9,11,13,15\}$$
$$B = \{2,3,5,7,11,13\}$$
$$C = \{4,5,6,8,9,11,12\}$$

a) $A \cup B =$

b) $(A \cap B) \cup C' =$

c) $A' \cap (B \cap C) =$

Bloque 7 | Técnicas de Conteo

7.1 Elementos básicos

Introducción

A las técnicas de conteo también se les conoce como análisis combinatorio; permite determinar el número posible de resultados lógicos que cabe esperar al realizar algún experimento o evento, sin necesidad de enumerarlos todos.

Con frecuencia se presentan problemas en los cuales, por ejemplo, una institución bancaria debe proporcionar a los usuarios una tarjeta de crédito o de débito; una compañía de teléfonos debe asignar a cada suscriptor un número; y un gobierno estatal una placa de circulación para cada vehículo. La solución de este tipo de problemas implica calcular cuántos subconjuntos distintos se pueden formar con un conjunto de números. El sistema elegido, debe ser lo suficientemente amplio para cubrir el número de usuarios previsto.

A cada número, objeto o suceso se le llama *elemento*; a cada colección o grupo de elementos se le identifica como una combinación, y a cada ordenamiento único se le identifica como una *permutación*.

Una **combinación** es un conjunto de elementos diferentes en cualquier orden.

Una **permutación** se caracteriza por el orden de los elementos que la forman.

Actividades de Apertura

Una cafetería indica que con los ingredientes lechuga, tomate, salsa de tomate y cebolla, se puede preparar una hamburguesa, en una de sus 16 formas posibles. ¿Está el anuncio en lo correcto?

Para corroborar la información del anuncio, indica con un Sí que se añade ese ingrediente, o de lo contrario, un No. Utiliza la siguiente tabla:

	Lechuga	Tomate	Salsa	Cebolla
1	Sí	Sí	Sí	Sí
2	Sí	Sí	Sí	No
3	Sí	Sí	No	Sí
4	Sí	Sí	No	No
5	Sí	No	Sí	Sí
6	Sí	No	Sí	No
7	Sí	No	No	Sí
8	Sí	No	No	No
9	No	Sí	Sí	Sí
10	No	Sí	Sí	No
11	No	Sí	No	Sí
12	No	Sí	No	No
13	No	No	Sí	Sí
14	No	No	Sí	No
15	No	No	No	Sí
16	No	No	No	No

Si con esos cuatro ingredientes puede prepararse la hamburguesa en 16 formas distintas, ¿de cuántas formas podría prepararse si hubiera 10 ingredientes disponibles? ¿Es necesario enumerarlos para saber de cuántas formas pueden prepararse una hamburguesa?

Actividades de Desarrollo

1. Revisa la imagen y responde lo que se plantea a continuación:

a) ¿De cuántas maneras distintas se pueden acomodar las prendas?

b) ¿De cuántas maneras distintas pueden combinarse las prendas si se las pone un compañero?

2. Supongamos que un restaurante ofrece 5 entradas, 4 platos principales y 3 postres. ¿De cuántas formas un cliente puede ordenar una comida?

3. ¿De cuántas maneras pueden sentarse 10 personas en una banca, si hay 4 espacios disponibles?

4. En una clase de 10 alumnos, se van a distribuir 10 premios. ¿De cuántas formas puede hacerse si los premios son diferentes?

5. Encuentra el número de distintos grupos de tres personas que se pueden formar, a partir de un grupo de 10 personas.

6. Calcula cuántos números enteros diferentes de tres dígitos se pueden formar con los dígitos 1, 2, 3, 4, 5 si no pueden repetirse.

7. Tu madre desea obsequiarte una laptop y una mochila para tu cumpleaños. Visitan a un distribuidor de equipos de cómputo; revisan cinco modelos de laptops (A, B, C, D, E) y 10 modelos de mochilas (1, 2, 3, 4, 5, 6, 7, 8, 9, 10). ¿De cuántas formas se puede elegir el par de artículos?

8. Si en casa tienes cuatro objetos para decorar cada una de las esquinas de tu sala; determina de cuántas maneras distintas se pueden acomodar los adornos, para que sean vistos.

Actividades de cierre

Diseñar un Muestrario en Papel Opalina utilizando los colores como elementos de insumo, 7 colores básicos (rojo, verde, amarillo, azul, rosado, naranja, negro).

El muestrario será conformado con agrupaciones No ordenadas de 3 colores, mismos que no deberán repetirse en grupos siguientes.

7.1.2. Espacio muestral, evento y diagrama de árbol

Introducción

El conjunto de todos los resultados posibles de un experimento aleatorio se conoce como **espacio muestral**; cada uno de ellos es un **punto muestral** y el resultado que obtenemos o esperamos obtener al realizar una o varias veces el mismo experimento es un **evento o suceso**.

Los **eventos** se clasifican en dos: **Evento elemental o simple**: Es conjunto formado por un solo resultado. **Evento compuesto:** Es el conjunto formado por dos o más posibles resultados**.** Si el experimento se repite dos o más veces, el número de resultados posibles cambia notoriamente al igual que la naturaleza de los puntos muestrales. Veamos.

Ejemplo: Determinamos el espacio muestral resultante de echar dos volados.

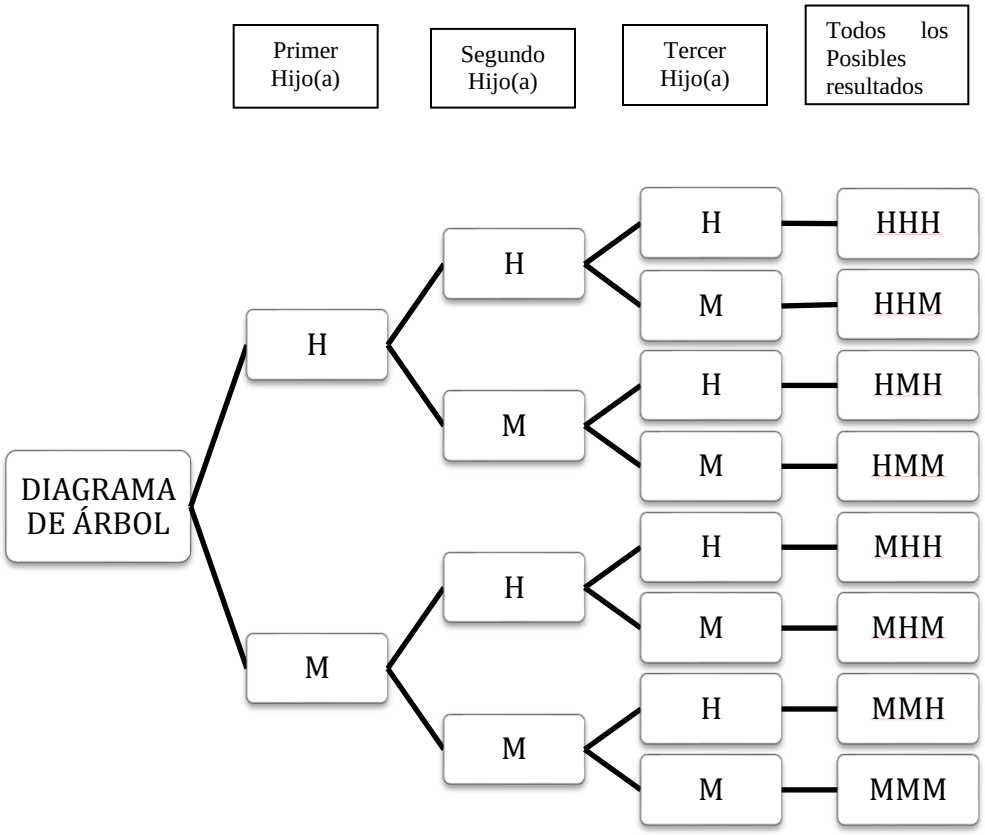
Solución: se trata de un experimento formado por dos repeticiones del mismo tipo: echar un volado. Pensando un poco, encontramos que los resultados posibles pueden ser 2 águilas, 1 águila y 1 sol, 1 sol y un águila, 2 soles. Usando las iníciales correspondientes, podemos anotar este espacio muestral como sigue:

(**AA, AS, SA, SS**)Como acabamos de ver, la determinación del espacio muestral de experimentos que implican una o dos repeticiones de su tipo, como echar uno o dos volados, tirar dos veces un dado, planear tener dos hijos, es muy fácil ya que basta una simple inspección. No obstante, cuando se quiere conocer a todos los resultados posibles de una serie de experimentos o repeticiones del mismo tipo, en el cual los puntos muestrales son necesariamente eventos conjuntos, esa facilidad desaparece rápidamente y se torna más y más difícil a medida que se alarga la serie. Por fortuna, para resolver estos casos existe una técnica conocida como:

DIAGRAMA DE ÁRBOL, cuya aplicación conduce metódicamente al espacio muestral que se quiere conocer. Explicaremos en que consiste recurriendo a un ejemplo.

Ejemplo; Se nos informa que una pareja planea procrear tres hijos. ¿De cuántas maneras diferentes puede suceder este hecho? En otras palabras: ¿Cuál es el espacio muestral del experimento "procrear 3 hijos"?

Solución: Sea H el evento "ser Hombre" y M el evento "ser Mujer". Es evidente que el primer hijo puede ser tanto de un sexo como del otro; las dos ramas primarias de la figura siguiente ilustran la situación. Para el segundo y el tercer hijo, la posibilidad de cualquiera de los dos sexos sigue siendo la misma; en consecuencia, de cada rama primaria se desprenden dos secundarias y, de cada secundaria, dos terciarias con lo cual se termina el diagrama necesario para resolver el problema.



7.2 Principio fundamental de la multiplicación

Introducción

El **principio fundamental** de **conteo** establece que si hay “p” formas de hacer una cosa, y “q” formas de hacer otra cosa, entonces hay “p × q” formas de hacer ambas cosas, (resultados posibles del experimento). El **principio** de **conteo** puede extenderse a situaciones donde tenga más de 2 opciones, por ejemplo, si hay “*p”* formas de hacer una cosa, “*q”* formas para una segunda cosa, y “*r”* formas de hacer una tercera cosa, entonces hay “$p \times q \times r$” formas de hacer las tres cosas.

Observa el video y analiza.
https://www.youtube.com/watch?v=u6qc-hglUcg

Ejemplo 1.

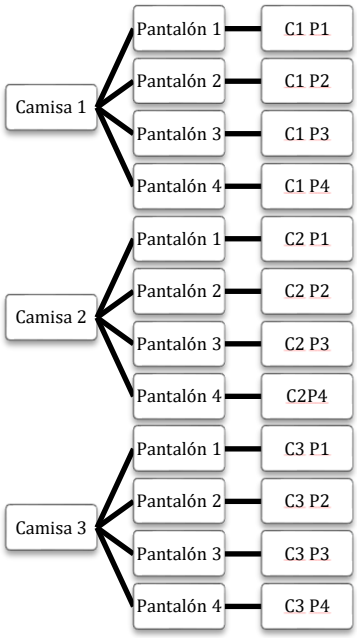
Silvia tiene 3 camisas (Blusas), llamémoslas C_1, C_2, C_3, y 4 pantalones, (llamémoslos P_1, P_2, P_3, P_4). Entonces Silvia puede hacer, 3 × 4 = 12 combinaciones posibles:

C_1P_1, C_1P_2, C_1P_3, C_1P_4

C_2P_1, C_2P_2, C_2P_3, C_2P_4

C_3P_1, C_3P_2, C_3P_3, C_3P_4

Esta ordenación puede ilustrarse con un “diagrama de árbol” como se muestra en la figura siguiente:



Ejemplo 2.

Suponga que lanza un dado de 6 lados y elige una ficha de domino de 28 piezas.

Solución. Hay 6 resultados posibles con el dado, y 28 resultados posibles con el grupo de piezas del domino, así, hay un total de 6 × 28 = 168 resultados posibles del experimento, ¿será práctico elaborar el diagrama de árbol para saber el número de resultados posibles?

Ejemplo 3.

Suponga que lanza un dado de 6 lados y elige una ficha de domino de 28 piezas, al mismo tiempo se tira un volado (se arroja una moneda al aire). ¿Cuántos resultados posibles existen?

Solución. Hay 6 resultados posibles con el dado, y 28 resultados posibles con el grupo de piezas del domino y dos resultados posibles para la moneda, así, hay un total de 6 × 28 x 2 = 336 resultados posibles del experimento, nuevamente la pregunta podría ser ¿será práctico elaborar el diagrama de árbol para saber el número de resultados posibles?

Ejemplo 4.

¿Cuántos números de 3 cifras distintas pueden formarse con los dígitos 1, 3, 5, 7,9?

Solución. El primer evento es seleccionar el primer dígito, el segundo evento es seleccionar el segundo dígito y el tercer evento es seleccionar el tercer dígito. Hay 5 maneras de seleccionar el primer dígito, y para cada manera de seleccionar el primero habrá 4 maneras de seleccionar el segundo dígito (pues de los 5 dígitos que disponemos, ya usamos 1 y nos quedan 4), y por cada una de esas maneras habrá 3 maneras de seleccionar el tercer dígito (pues de los 5 dígitos que disponemos, ya usamos 2 y nos quedan 3), entonces, hay 5 x 4 x 3 = 60 números que podemos formar.

En los ejemplos anteriores del principio de multiplicación, del enunciado se deducen fácilmente los eventos y la cantidad de resultados posibles para cada uno, lo que hace que apliquemos el principio de multiplicación directamente, casi con los ojos cerrados y como si fuera una fórmula. Sin embargo, habrá situaciones un poco más complejas donde el principio de la multiplicación no se aplique directamente. Veamos el siguiente ejemplo:

Ejemplo 5.

¿Cuántos números capicúas de 5 dígitos hay? Nota: Un número capicúa es aquel que se lee igual de izquierda a derecha que de derecha a izquierda, por ejemplo 12321 y 72227.

Solución. Como se puede observar, el enunciado no nos habla de eventos que se realizan, ni mucho menos de la cantidad de posibles resultados para cada evento. Lo primero será razonar un poco sobre el ejercicio y ver a dónde nos lleva: Si tenemos un número de 5 cifras que se lee igual de izquierda a derecha que de derecha a izquierda entonces el 1er y 5to dígito deben ser iguales, de la misma forma el 2do y 4to dígito son iguales, y el 3er dígito puede ser lo que sea (pues ocupa la misma posición de izquierda a derecha que de derecha a izquierda); entonces basta con elegir los primeros tres dígitos de nuestro número (y el 4to y 5to dígito los igualamos al 2do y 1er dígito respectivamente) para formar un número capicúa de 5 dígitos.

Este razonamiento ya nos deja ver los eventos que debemos considerar: el primer evento es elegir el 1er dígito, el segundo evento es elegir el 2do dígito, y el tercer evento es elegir el 3er dígito. Hay 9 formas de elegir el primer dígito (los dígitos del 1 al 9, el 0 no, porque no sería un número de 5 dígitos, por ejemplo 02620 no es válido), y por cada una de estas formas hay 10 formas de elegir el 2do dígito (los dígitos del 0 al 9), y por cada una de estas hay 10 formas de elegir el 3er dígito (los dígitos del 0 al 9), así, hay un total de 9 x 10 x 10 = 900 números capicúas de 5 dígitos.

Es importante notar lo siguiente: para poder hacer uso del principio de multiplicación es necesario que de los eventos que estemos considerando, para cada uno de los resultados de un evento, la cantidad de opciones del siguiente evento sea la misma. Por ejemplo, en el del ejemplo 1, la selección de la camisa no altera la cantidad de opciones de pantalón, pues para cada camisa de entre las 3 que hay, siempre habrá 4 posibles pantalones, haciendo entonces que haya 4+4+4=3 x 4 = 12 combinaciones.

¿Qué pasa entonces si, en el Ejemplo 1, añadimos la restricción de que la primera camisa no se ve bien con 2 de los 4 pantalones y que por lo tanto no debemos considerar esas combinaciones? No podemos usar solo el principio de multiplicación, pues si la camisa seleccionada es la primera entonces hay 2 opciones de pantalón, y si la camisa seleccionada es cualquiera de las otras 2 entonces hay 4 opciones de pantalón, así, habrá en total 1x2 + 2x4 = 10 combinaciones. Lo que hicimos entonces fue dividir en dos casos el problema, cuando la primer camisa era seleccionada y cuando no, y para cada caso resolvimos el total de combinaciones usando el principio de multiplicación, y después usamos el **principio de la suma** para llegar a que la respuesta es 1x2 + 2x4 = 10**.**

A continuación, presentamos de forma detallada el principio de la suma.

7.3 Principio fundamental de la suma

Observa el video y analiza.
https://www.youtube.com/watch?v=u6qc-hglUcg

Si un evento «A» se puede realizar de «m» maneras diferentes, y otro evento «B» se puede realizar de «n» maneras diferentes, además, ***si ocurre uno no puede ocurrir el otro***, (son mutuamente excluyentes) entonces, el evento A o el evento B, se realizarán de m + n formas.
Es decir, aquí ocurre A u ocurre B. **El «o» indica suma.**

Ejemplo 6: ¿De cuántas formas se puede ir de la ciudad “A” hasta la ciudad “B” sabiendo que se dispone de 3 autos diferentes y 4 trenes distintos? El viaje se puede realizar en auto o en tren.

Solución: Dado que el viaje se puede realizar en auto o en tren, se tienen 3 + 4 = 7 opciones diferentes para ir de la ciudad “A” hasta la ciudad “B”.

Ejemplo 7:¿Cuántos resultados se pueden obtener si se lanza un dado o una moneda?

Solución: Es evidente que al lanzar un dado se tienen seis resultados posibles, y al lanzar una moneda, se tienen 2 resultados posibles, por lo tanto:

El número de resultados posibles es *2 + 6 = 8*

Ejemplo 8: ¿Cuántos números pares de 2 cifras empiezan con 1 o 2?

Nota: el cero se considera par

Solución: Los que inician con uno son (12, 14, 16, 18, 10) más los que inician con 2 son, (22, 24, 26, 28, 20) en total son 10 números.

En los ejemplos anteriores, se deduce fácilmente cuáles son los eventos y de cuántas formas pueden ocurrir cada uno. Sin embargo, habrá situaciones más complejas donde seremos nosotros quienes determinamos los eventos a considerar, los cuales deben cumplir con que son excluyentes y que dichos eventos cubran todos los posibles casos. Veamos el siguiente ejemplo:

Ejemplo 9:

Eliseo planea ir de la ciudad A a la ciudad B y de ahí a la ciudad C, cada uno de los 2 viajes puede hacerlo en camión, tren, auto o avión, pero con la condición de que al menos 1 de los viajes sea en avión, para no cansarse tanto. ¿De cuántas formas puede hacerlo?

Solución:

Podemos dividir en dos eventos: Eliseo viaja de A a B en avión, o no viaja de A a B en avión. Si se va de A a B en avión, entonces hay 4 formas de completar su viaje (avión-tren, avión-camión, avión-auto, avión-avión), y si no se va de A a B en avión entonces hay 3 formas de completar su viaje (tren-avión, camión-avión, auto-avión), en total son 4+3=7 formas de completar su viaje. Nótese que aquí fuimos nosotros quienes determinamos los eventos a considerar, los cuales son excluyentes y cubren todos los posibles casos

En muchas ocasiones, como se mencionó al final de la sección del principio de multiplicación con un ejemplo, habrá problemas en los que ambos principios sean utilizados en conjunto para poder resolverlos, veamos otro ejemplo:

Ejemplo 10

Eliseo planea un nuevo viaje, ahora irá de la ciudad A a la ciudad B, luego a la ciudad C y finalmente a la ciudad D, puede hacerlo en camión, tren, auto o avión, pero con la condición de que a lo más 1 de los 3 viajes sea en avión, pues ya no tiene tanto dinero. ¿De cuántas maneras puede hacerlo?

Solución:

Como tenemos una restricción, primero hay que dividir en eventos excluyentes que juntos consideren todos los posibles casos (para después sumarlos usando el principio de la suma), y para cada uno de estos eventos encontraremos el total de posibles viajes usando el principio de multiplicación. O bien ningún viaje es en avión, o bien uno de ellos es en avión, entonces los eventos que consideramos son los siguientes: el primero es que ninguno de los viajes se hace en avión, el segundo es que solo el viaje de A a B se hace en avión, el tercero es que solo el viaje de B a C es en avión, y el cuarto es que solo el viaje de C a D se hace en avión. Si ningún viaje es en avión, cada uno de los 3 viajes se hace en camión, tren o auto y usando el principio de multiplicación hay 3x3x3=27 formas de hacerlo. Si se viaja de A a B en avión, entonces hay 3 formas de viajar de B a C y 3 formas de viajar de C a D, es decir, 3x3=9 formas. Análogamente, si solo se viaja en avión de B a C hay 3x3=9 formas, y si solo se viaja en avión de C a D hay 3x3=9 formas. Ahora, por el principio de la suma, las maneras totales en que Eliseo puede viajar son 27+9+9+9=54 formas.

7.3.1 Notación factorial

Sea *n,* un número entero y positivo, entonces el factorial de *n,* que se simboliza *n!* y se lee "*n* factorial", es el producto de todos los enteros positivos desde 1 hasta *n.* Así.

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120.$$

Obsérvese que $5! = 5 \times 4!$

$$= 5 \times 4 \times 3!$$
$$= 5 \times 4 \times 3 \times 2!$$
$$= 5 \times 4 \times 3 \times 2 \times 1!$$

Por definición: $0! = 1! = 1$

7.3.1.1 operaciones con factoriales, combinaciones y permutaciones

Actividades de Apertura

Definimos una combinación, como un acomodo de "n" objetos diferentes, en subconjuntos de "r" elementos, (con n ≥ r) en donde el orden en que están dispuestos los objetos, no importa.

La notación factorial es útil cuando se trata de calcular el número de posibles combinaciones o permutaciones para un evento dado.

Ejemplo 11: Se tienen 3 colores diferentes (por ejemplo, Verde, Amarillo y Café) y se desea saber ¿cuántas combinaciones diferentes de 2 colores son posibles?

Solución: Las combinaciones se pueden calcular "a píe", del siguiente modo Verde-Amarillo, Verde-Café, Amarillo-Café, son tres posibilidades, es muy importante hacer notar que este resultado se podría calcular utilizando la siguiente fórmula para combinaciones:

$$ {}_nC_r = \frac{n\ !}{r\ !\ (n-r)\ !} $$

Ejemplo 12. Seleccionar dos platillos de verduras de un menú en el que hay que cinco opciones diferentes.

Solución: ${}_5C_2 = \frac{5\ !}{2\ !\ (5-2)\ !} = \frac{5\ x\ 4\ x\ 3\ !}{2\ !\ (3)\ !} = 10$

Ejemplo 13. De un grupo de 9 personas, se van a seleccionar 3, para experimentar una nueva vacuna contra el coronavirus, ¿Cuántos grupos diferentes de tres personas se pueden formar?

Solución: ${}_9C_3 = \frac{9\ !}{3\ !\ (9-3)\ !}\ \frac{9\ x\ 8\ x\ 7\ x\ 6\ !}{3\ !\ (6)\ !} = \frac{9\ x\ 8\ x\ 7}{3\ x\ 2\ x\ 1} = 84$

7.4 Combinaciones y Permutaciones

Observa el video y analiza.
https://www.youtube.com/watch?v=DhOeAPRXGxM

Actividades de Desarrollo

Analiza los siguientes ejemplos:

Esta fórmula sirve para calcular el número de combinaciones con un total de "n" objetos en subgrupos de "r" y se lee del siguiente modo "n en r" y se simboliza de la siguiente forma:

$$ \binom{n}{r} = \frac{n\ !}{r\ !\ (n-r)\ !} $$

para el ejemplo anterior n = 3 colores y r = 2, (grupos de dos en dos).

Utilizando la fórmula ${}_3C_2 = \frac{3\ !}{2\ !\ (3-2)\ !} = \frac{3\ x\ 2\ x\ 1}{2\ x\ 1\ (1)} = \frac{6}{2} = 3$

Es probable que hasta este punto pueda parecer trivial esta fórmula, sin embargo, resulta de extraordinaria utilidad cuando se trata de resolver situaciones como la siguiente:

Ejemplo 14: ¿Cuántas combinaciones de tres colores diferentes se pueden formar, si se tienen 10 colores disponibles?

Solución: Utilizando nuestra fórmula $_nC_r = \frac{n\,!}{r\,!\,(n-r)\,!} =$ $_{10}C_3 = \frac{10\,!}{3\,!\,(10-3)\,!} =$

$= \frac{10\ x\ 9\ x\ 8\ x\ 7\,!}{3\,!\,(10-3)\,!} =$ $= \frac{10\ x\ 9\ x\ 8\ x\ 7\,!}{3\,!\,(7)\,!}$ cancelando términos 7! del numerador con 7! del denominador, $= \frac{10\ x\ 9\ x\ 8}{3\,!} = \frac{10\ x\ 9\ x\ 8}{3\ x\ 2\ x\ 1}$ = 120 combinaciones posibles

Actividades de cierre

Resuelva los siguientes problemas propuestos:

1. Con las 5 vocales, ¿cuántos grupos de 3 letras podemos formar?, teniendo en cuenta que ninguna letra se puede repetir y que el orden no importa.

2. De un grupo de 10 alumnos se seleccionan 5 para jugar en el equipo de basquetbol ¿Cuántos grupos diferentes se pueden formar? (considere que todos pueden jugar en cualquier posición).

3. En una final de futbol se seleccionan 5 jugadores de un equipo para el lanzamiento de penaltis. ¿Cuántos grupos diferentes se podrían formar?

4. En una carrera de caballos con 9 participantes tienes que elegir a los 3 caballos ganadores (no importa el orden de llegada). ¿Cuántos posibles resultados podrían darse?

5. Calcular el número de combinaciones de 8 elementos tomados de 4 en 4.

6. En una clase de 50 alumnos se quiere elegir un comité formado por tres alumnos. ¿Cuántos comités diferentes se pueden formar?

7. En una bodega hay siete tipos diferentes de vinos. ¿De cuántas formas se pueden elegir cuatro?

8. ¿Cuántas diagonales tiene un hexágono y cuántos triángulos se pueden formar con sus vértices?

9. A una fiesta asisten 12 personas y se intercambian saludos entre todos. ¿Cuántos saludos se han intercambiado?

7.4.1 Permutaciones

Actividades de Apertura

Supón que tienes tres cifras diferentes, por ejemplo 1, 2, 3, ¿Cuántos números diferentes de dos cifras se pueden formar, sin que haya dos cifras iguales? La respuesta se puede ilustrar como sigue, 12, 13, 23, 21, 31, 32, claramente se entiende que 12 o 21, son dos números diferentes y cuenta como dos posibilidades distintas, esa es la diferencia respecto a las combinaciones, en combinaciones el orden no es relevante, en permutaciones el orden si es importante.

El número de permutaciones de un subconjunto de "r" elementos seleccionados de un conjunto de "n" elementos diferentes, está dado por: ${}_nP_r = \frac{n\,!}{(n-r)\,!}$ algunos autores la expresan

$$P_r^n = \frac{n\,!}{(n-r)\,!}$$

Si deseamos utilizar la fórmula para el ejemplo anterior

$${}_3P_2 = \frac{3\,!}{(3-2)\,!} = \frac{3\,!}{1\,!} = 3 \text{ x } 2 \text{ x } 1 = 6$$

Actividades de Desarrollo

Analiza el siguiente ejemplo:

Ejemplo 15: ¿De cuántas formas diferentes es posible agrupar las cifras 1, 2, 3, en grupos de tres?

Solución: 123, 132, 213, 231, 312, 321.

Lo anterior también podría enunciarse como ¿Cuántos números diferentes de tres cifras pueden formarse con los números 1, 2 y 3?

Utilizando la fórmula ${}_3P_3 = \frac{3\,!}{(3-3)\,!} = \frac{3\,!}{0\,!} = \frac{3\,!}{1} = 3 \text{ x } 2 \text{ x } 1 = 6$

Recuerda que el factorial de cero se define igual a 1 (0! = 1)

Actividades de cierre

Resuelva los siguientes problemas propuestos:

1. ¿Cuántas banderas diferentes, de tres franjas horizontales de igual ancho y de colores distintos, pueden confeccionarse a partir de siete colores diferentes?

2. Una tienda de pinturas requiere diseñar muestrarios de los colores que vende, colocándolos en una fila de 5 lugares; si tienen 4 verdes diferentes, 5 amarillos diferentes, 2 azules diferentes y 3 rosados diferentes, calcula el número de muestrarios diferentes que se pueden diseñar si:
   a) No hay restricción alguna
   b) El primer color debe ser azul
   c) El primero y el último deben ser verdes

3. Con todas las letras de la palabra DISCO, ¿Cuántas “palabras” distintas se pueden formar?

4. Un grupo de cuatro matrimonios asiste al teatro, calcula cuántas son las maneras como pueden sentarse en una fila si:
   a) No existe restricción alguna
   b) Los y las mujeres deben quedar alternados

5. La directiva de una asociación civil deberá tomarse una fotografía en la cual aparezcan sentados en fila, sus cinco directivos. ¿De cuántas maneras diferentes pueden sentarse los cinco directivos?

6. El gobierno de un Estado de la Federación, decide cambiar las placas de circulación de los automóviles particulares. Para ello, se considera incluir dos letras distintas del alfabeto, seguidas de tres dígitos distintos del 0 al 9 inclusive. ¿Cuántas placas podrían fabricarse con estas características?

7. Un funcionario de un Banco decide que los números de las tarjetas de crédito se cambien, de manera que no se repitan las letras o los números de cada una; mismas que incluirán dos letras del alfabeto, seguidas de cuatro dígitos. Calcula el número de tarjetas que se podrían fabricar.

8. Calcula el número de permutaciones diferentes que se pueden formar con las letras A, B, C, D, E, F que contengan 3 letras cada una.

9. Para tu clase de Literatura debes leer 3 libros en el semestre. Si tienes en casa 12 libros de distintos autores que te pueden interesar, determina de cuántas formas se pueden seleccionar los 3 libros.

10. Cuatro libros distintos de matemáticas, seis diferentes de física y dos diferentes de química se colocan en un estante. De cuántas formas distintas es posible ordenarlos si:
    a) Los libros de cada asignatura deben estar todos juntos.
    b) Solamente los libros de matemáticas deben estar juntos

7.5 Teorema de Bayes

Introducción

En la vida cotidiana, tomamos muchas decisiones basándonos en información de conocidos, medios de comunicación, entre otros. Cuando decidimos que queremos comprar un teléfono celular, una motocicleta u otro artículo, lo hacemos considerando la fiabilidad, durabilidad o rendimiento de una marca. En la mayoría de casos desconocemos que estos artículos se producen en diferentes plantas en el mundo, cada instalación tiene sus propias estadísticas de calidad.

Cuando nuestro papá decide comprar un auto Toyota Corolla 2012, porque su hermano en Estados Unidos tiene uno y es excelente, no sabe que el auto que él comprará fue fabricado en Brasil, con especificaciones distintas al de su hermano fabricado en Canadá.

El Teorema de Bayes trata de establecer la probabilidad de que ocurra cierto evento (Auto sin fallas), considerando la ocurrencia de otro evento previo (Fabricado en Canadá o Brasil). El análisis y cálculo de estas probabilidades nos permiten tomar mejores decisiones.

Actividades de Apertura

Imagina que para recabar fondos y pagar la graduación, un grupo de alumnos decide, hornear y vender mantecadas. Fátima pertenece a este grupo y sabe mucho de pastelería, pero no puede hacer todos los panes, por lo que algunos de sus compañeros se ofrecen a cocinarlos en su casa con la receta y recomendaciones de Fátima. Como es lógico, no a todos los alumnos les saldrán mantecadas perfectas. Si te dieran a escoger ¿Comprarías una de las mantecadas de Fátima o de otros de sus compañeros?, ¿Crees que el porcentaje de mantecadas casi perfectas será igual para todos los alumnos, aunque sea la misma receta? Socializa las opiniones y arguméntalas.

Teorema de Bayes

Las particiones son varios eventos A_1, A_2…, A_n, que forman un espacio muestral S, el teorema de Bayes combina al evento S con otro evento B por medio de la probabilidad condicional quedándonos la siguiente formula.

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots P(A_n)P(B|A_n)}$$

Observa el video y analiza.
https://youtu.be/c0r5-_52fwM

Ejemplo 1: Tres plantas de autos A, B y C producen respectivamente 50%, 30% y 20% del número total de autos de una marca, los porcentajes de desperfectos en estas plantas son 3%, 4% y 5%, si se selecciona al azar un auto, hallar la probabilidad de: (a) que el auto sea defectuoso de la planta A y (b) este mismo auto no sea defectuoso de la planta C.

Solución:

El 50% de autos producidos en la planta A se escribe en forma decimal como 0.5, y el porcentaje de autos con desperfectos de esa planta es de 3%, o sea 0.03. Por lo anterior, el producto (0.5*0.03) hace referencia a la planta A. para las demás plantas es similar.

Autos defectuosos de la planta A.

$$a)\ P(A|B) = \frac{0.5 * 0.03}{(0.5 * 0.03) + (0.3 * 0.04) + (0.2 * 0.05)} = \frac{0.015}{0.037} = 40.54\%$$

Cuando nos solicitan el porcentaje de autos SIN desperfectos, debemos usar el complemento para llegar al 100%, es decir, si la planta C produce un 5% de autos defectuosos, producirá un 95% de autos NO defectuosos. Por lo anterior, para la planta C usaremos (0.2*0.95). Para las demás plantas es similar el proceso.

Autos NO defectuosos de la planta C.

$$b)\ P(A|B) = \frac{0.2 * 0.95}{(0.5 * 0.97) + (0.3 * 0.96) + (0.2 * 0.95)} = \frac{0.19}{0.963} = 19.73\%$$

Ejemplo 2: En cierta escuela 4% de los hombres y 1% de las mujeres miden 1.80 metros o más, además 40% de los estudiantes son hombres, ahora bien, si se selecciona un estudiante al azar y mide más de 1.80 m, ¿cuál es la probabilidad de que sea mujer?

Solución: El 1% de las mujeres miden 1.80 m o más, por lo que se escribe en forma decimal como 0.01, si el 40% de los estudiantes son hombres, el 60% del estudiantado será mujer, es decir trabajaremos con el producto (0.6*0.01) para hacer referencia a las mujeres que cumplan las condiciones señaladas, para los hombres es similar.

Mujeres que miden 1.80 m o más.

$$P = (A|B) = \frac{0.6 * 0.01}{(0.4 * 0.04) + (0.6 * 0.01)} = \frac{0.006}{0.022} = 27.27\%$$

Actividades de cierre

Cuando vayamos a realizar compras de artículos diversos, deberíamos incorporar varios criterios como son: calidad, relación costo-beneficio, durabilidad, pertinencia, uso que le daremos, opinión de expertos, etc.

Entre mayor sea el costo del artículo que vamos a adquirir o seleccionar, mayor deberá ser la cantidad de tiempo invertido en la investigación y valoración de las características del producto y sus posibles alternativas.

Existen muchas agencias gubernamentales e independientes que hacen estudios de calidad de variados productos, pero no sirven de mucho si nosotros no las consultamos. Algunos ejemplos:

EJERCICIO 1: Un equipo de fútbol se conforma de cuatro líneas de jugadores: portero (9%), defensas (37%), medios (27%) y delanteros (27%), donde los goles anotados por el equipo en cada línea son: portero 1 gol, defensas 11 goles, medios 15 goles y delanteros 23. Supongamos que en un partido se gana por un gol. ¿Qué probabilidad hay de que éste fuera anotado (a) por la defensa, (b) por los medios y (c) por los delanteros?

Total de goles = ____ + ____ + ____ + ____ = ____

% goles anotados por: Portero = ____ / ____ = _____%

Defensas = ____ / ____ = _____%

Medios = ____ / ____ = _____%

Delanteros = ____ / ____ = _____%

Probabilidad de que un gol fuera anotado por LA DEFENSA

$a)\ P(A|B) =$ ——————————————— $=$ %

Probabilidad de que un gol fuera anotado por LOS MEDIOS

$b)\ P(A|B) =$ ——————————————— $=$ %

Probabilidad de que un gol fuera anotado por LOS DELANTEROS

$c)\ P(A|B) =$ ——————————————— $=$ %

EJERCICIO 2: A un congreso asisten 100 personas de las cuales 65 son hombres y 35 mujeres, se sabe que el 10% de los hombres y el 6% de las mujeres tienen estudios a nivel maestría. Si se selecciona un maestro al azar, ¿qué probabilidad existe de que sea (a) mujer y (b) hombre?

Probabilidad de que un asistente con grado de maestría sea MUJER

$a)\ P(A|B) =$ ——————————————— $=$ %

Probabilidad de que un asistente con grado de maestría sea HOMBRE

$b)\ P(A|B) =$ ——————————————— $=$ %

Fuentes consultadas

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2008). *Estadística aplicada a los negocios y a la economía.* Mc Graw Hill.

Rivera Castillo, M. (2014). *Probabilidad y estadística.* Gafra.

Márquez, E.M.A. (2007). Probabilidad y Estadística. SEP. CFE. COLECCIÓN DGETI. México D.F.

Garza Olvera, B. (2014). Estadística y Probabilidad. Ed. Pearseon.

Directorio

Dr. Rafael Sánchez Andrade
Jefe de la Dirección General de Educación Tecnológica Industrial y de Servicios

Ing. Luis Miguel Rodríguez Barquet
Director Académico de Innovación Educativa

Mtra. Laura Leal Sorcia
Subdirectora de Innovación Académica

MC Gerardo Valdés Bermudes
Presidente de la Academia Nacional de Matemáticas de la DGETI

MC Luis Manuel Guerra Franco
Secretario de la Academia Nacional de Matemáticas de la DGETI

MC Óscar Villalpando Barragán
Coordinador de la Mesa de trabajo de Probabilidad y Estadística

ME Omar Eduardo De la Torre Aldama
Edición de la obra

MC Gerardo Valdés Bermudes
Edición de la obra

Ing. Norma Patricia Hernández Tamez
Edición de la obra

Academia Nacional de Matemáticas

Integrantes de la Academia Nacional de Matemáticas que participaron en la elaboración de la obra.

Nombre	Plantel	Estado
Juan Carlos Díaz Puga	CBTIS 39	Aguascalientes
José Antonio Hirata Moyeda	CBTIS 140	Baja California
José Luis Colorado Betanzos	CBTIS 69	Baja California Sur
Ana Maria García Zuñiga	CETIS 2	Ciudad de México
Loan Alejandra Servin Rodríguez	CETIS 52	Ciudad de México
Brillante Zavala Centeno	UAC	Campeche
Yudibeth Sánchez Castellanos	CETIS 138	Chiapas
Miguel Angel Peña Ogaz	CBTIS 228	Chihuahua
Omar Eduardo De la Torre Aldama	CETIS 83	Coahuila
Felipe de Jesús Navarro Camacho	CBTIS 157	Colima
J. Armando Quezada López	CBTIS 89	Durango
Marcos Belisario González Loria	CBTIS 160	Estado de México
David Fernando López López	CBTIS 172	Guanajuato
Jesús Eugenio Ruiz Flores	CBTIS 60	Guanajuato
Eduardo Tomás Torres	CBTIS 178	Guerrero
Emilio Jaime Mendoza Gómez	CBTIS 199	Hidalgo
Eliseo Santoyo Teyes	CBTIS 226	Jalisco
Oscar Villalpando Barragán	CBTIS 12	Michoacán
Andrea Casillas Macías	CBTIS 94	Michoacán
Luis Manuel Guerra Franco	CBTIS 76	Morelos
Lucía Sánchez Ramos	CBTIS 74	Nuevo León
Eva Cruz Brena	CBTIS 183	Oaxaca
Julio Alberto González Negrete	CBTIS 86	Puebla
Gilmer de Jesús Pat Sánchez	CBTIS 111	Quintana Roo
Gerardo Valdés Bermudes	CBTIS 224	Sinaloa
Martín Vega Gómez	CETIS 128	Sonora
Eva María Sánchez Luna	CBTIS 249	Tabasco
Norma Patricia Hernández Tamez	CBTIS 007	Tamaulipas
Miguel Constantino Hernández Pérez	CETIS 132	Tlaxcala
Miguel Ángel Pavón Cordero	CBTIS 48	Veracruz
Silvia Leonor Martínez Quijano	CBTIS 80	Yucatán
Efraín Reyes Cumplido	CBTIS 104	Zacatecas

[i] MARQUEZ, E.M.A.(2007).Probabilidad y Estadística. SEP.CFE.COLECCION DGETI. México D.F.
[ii] Márquez Elías, M.A.(2007). Probabilidad y Estadística. DGETI.SEP.

MATEMÁTICAS

PROBABILIDAD Y ESTADÍSTICA

TERCERA EDICIÓN

SAMUEL FUENLABRADA

Mc Graw Hill

INCLUYE

Probabilidad y estadística

Tercera edición

Samuel Fuenlabrada de la Vega Trucíos
Instituto Politécnico Nacional

Revisores técnicos

Irma Fuenlabrada Velázquez
Departamento de Investigaciones Educativas
Centro de Investigación y de Estudios Avanzados
Instituto Politécnico Nacional

Bertha Vivanco Ocampo
Departamento de Investigaciones Educativas
Centro de Investigación y de Estudios Avanzados
Instituto Politécnico Nacional

Leandro Brito Barrera
Maestro en Ciencias en Ingeniería Mecánica
Escuela Superior de Ingeniería Mecánica y Eléctrica
Instituto Politécnico Nacional

México • Bogotá • Buenos Aires • Caracas • Guatemala • Lisboa • Madrid • Nueva York • San Juan • Santiago • Auckland • Londres • Milán • Montreal • Nueva Delhi • San Francisco • Singapur • St. Louis • Sydney • Toronto

Publisher de la división escolar: Jorge Rodríguez Hernández

Director editorial: Ricardo Martín del Campo

Editora de desarrollo: Talía Delgadillo Santoyo

Supervisora de producción: Jacqueline Brieño Álvarez

Diseño de portada e interiores: Código X, S. C.

Formación tipográfica: Overprint, S. A. de C. V.

Probabilidad y estadística

Tercera edición

Prohibida la reproducción total o parcial de esta obra,
por cualquier medio, sin la autorización escrita del editor.

DERECHOS RESERVADOS © 2008, respecto a la tercera edición por:
McGRAW-HILL / INTERAMERICANA EDITORES S. A. DE C. V.
A Subsidiary of The McGraw-Hill Companies, Inc.

Punta Santa Fe
Prolongación Paseo de la Reforma 1015
Torre A, piso 17, Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón
C.P. 01376, México, D.F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN 970-10-6229-9
ISBN 970-10-4703-6 (Segunda edición)

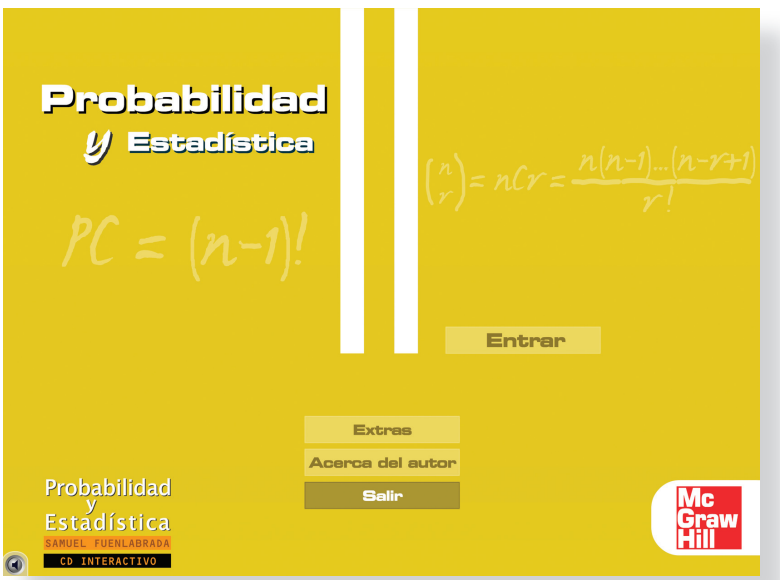
1234567890 09865432107
Impreso en México Printed in Mexico

Conoce tu CD

Uno de los valores agregados de esta nueva edición es el CD que acompaña a tu libro de texto. En este disco podrás encontrar evaluaciones y ejercicios adicionales.

Todos estos recursos harán que el aprendizaje de la disciplina sea más dinámico y atractivo.

¡Te invitamos a que pongas a prueba tus conocimientos!



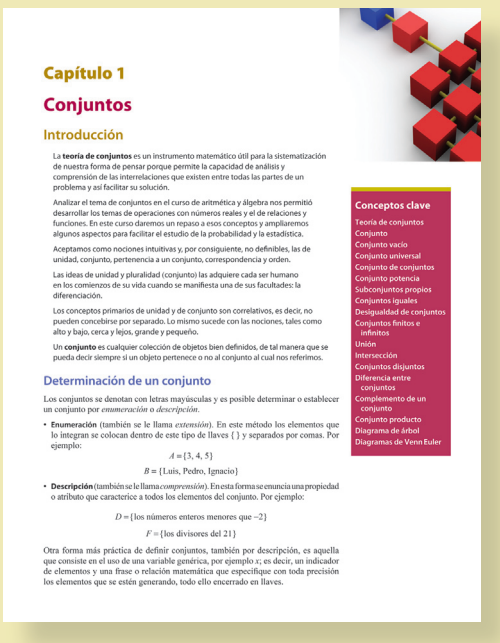
No necesitas tener instalado ningún programa en particular porque el software es autoejecutable y eres tú el que decide qué capítulos revisar y sobre todo, qué actividades realizar.

Toda la información está catalogada por capítulos y tienes la opción de imprimir tus evaluaciones para que puedas consultar con tu profesor cualquier duda.

Conoce tu libro

Organización

Para esta nueva edición, hemos mejorado la presentación de los temas para mejor referencia de profesores y alumnos. Este nuevo formato te permitirá ubicar con mayor facilidad las partes y secciones en las que se divide tu libro.



Capítulo 1

Conjuntos

Introducción

La **teoría de conjuntos** es un instrumento matemático útil para la sistematización de nuestra forma de pensar porque permite la capacidad de análisis y comprensión de las interrelaciones que existen entre todas las partes de un problema y así facilitar su solución.

Analizar el tema de conjuntos en el curso de aritmética y álgebra nos permitió desarrollar los temas de operaciones con números reales y el de relaciones y funciones. En este curso daremos un repaso a esos conceptos y ampliaremos algunos aspectos para facilitar el estudio de la probabilidad y la estadística.

Aceptamos como nociones intuitivas y, por consiguiente, no definibles, las de unidad, conjunto, pertenencia a un conjunto, correspondencia y orden.

Las ideas de unidad y pluralidad (conjunto) las adquiere cada ser humano en los comienzos de su vida cuando se manifiesta una de sus facultades: la diferenciación.

Los conceptos primarios de unidad y de conjunto son correlativos, es decir, no pueden concebirse por separado. Lo mismo sucede con las nociones, tales como alto y bajo, cerca y lejos, grande y pequeño.

Un **conjunto** es cualquier colección de objetos bien definidos, de tal manera que se pueda decir siempre si un objeto pertenece o no al conjunto al cual nos referimos.

Determinación de un conjunto

Los conjuntos se denotan con letras mayúsculas y es posible determinar o establecer un conjunto por *enumeración* o *descripción*.

- **Enumeración** (también se le llama *extensión*). En este método los elementos que lo integran se colocan dentro de este tipo de llaves { } y separados por comas. Por ejemplo:

$$A = \{3, 4, 5\}$$

$$B = \{\text{Luis, Pedro, Ignacio}\}$$

- **Descripción** (también se le llama *comprensión*). En esta forma se enuncia una propiedad o atributo que caracterice a todos los elementos del conjunto. Por ejemplo:

$$D = \{\text{los números enteros menores que } -2\}$$

$$F = \{\text{los divisores del 21}\}$$

Otra forma más práctica de definir conjuntos, también por descripción, es aquella que consiste en el uso de una variable genérica, por ejemplo x; es decir, un indicador de elementos y una frase o relación matemática que especifique con toda precisión los elementos que se estén generando, todo ello encerrado en llaves.

Conceptos clave

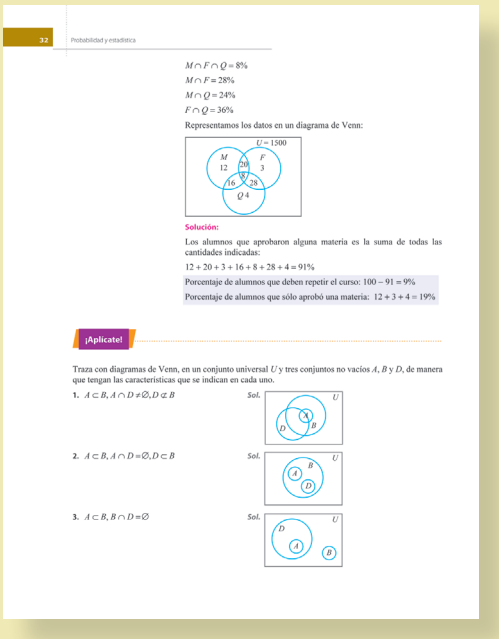
- Teoría de conjuntos
- Conjunto
- Conjunto vacío
- Conjunto universal
- Conjunto de conjuntos
- Conjunto potencia
- Subconjuntos propios
- Conjuntos iguales
- Desigualdad de conjuntos
- Conjuntos finitos e infinitos
- Unión
- Intersección
- Conjuntos disjuntos
- Diferencia entre conjuntos
- Complemento de un conjunto
- Conjunto producto
- Diagrama de árbol
- Diagramas de Venn Euler

Conceptos clave

En cada entrada de capítulo podrás ubicar los términos más importantes que se analizarán y que es importante memorices para continuar con tu progreso de aprendizaje. Estos términos representan la base que te permitirá adquirir conocimientos más complejos.

¡Aplícate!

Nueva sección de ejercicios que aparece después de haber estudiado un tema de extensión y complejidad considerable. Si tienes la capacidad de resolver los ejercicios ahí sugeridos, significa que tienes la capacidad para continuar con el resto de los temas del capítulo.



32 Probabilidad y estadística

$M \cap F \cap Q = 8\%$

$M \cap F = 28\%$

$M \cap Q = 24\%$

$F \cap Q = 36\%$

Representamos los datos en un diagrama de Venn:

Solución:

Los alumnos que aprobaron alguna materia es la suma de todas las cantidades indicadas:

$12 + 20 + 3 + 16 + 8 + 28 + 4 = 91\%$

Porcentaje de alumnos que deben repetir el curso: $100 - 91 = 9\%$

Porcentaje de alumnos que sólo aprobó una materia: $12 + 3 + 4 = 19\%$

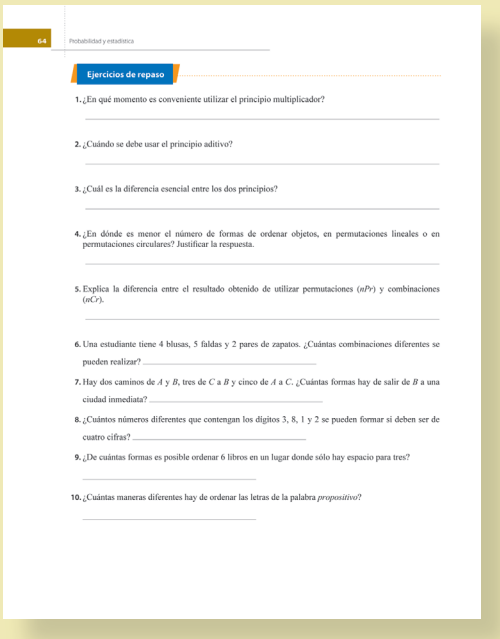
¡Aplícate!

Traza con diagramas de Venn, en un conjunto universal U y tres conjuntos no vacíos A, B y D, de manera que tengan las características que se indican en cada uno.

1. $A \subset B, A \cap D \neq \varnothing, D \not\subset B$ Sol.

2. $A \subset B, A \cap D = \varnothing, D \subset B$ Sol.

3. $A \subset B, B \cap D = \varnothing$ Sol.



64 Probabilidad y estadística

Ejercicios de repaso

1. ¿En qué momento es conveniente utilizar el principio multiplicador?

2. ¿Cuándo se debe usar el principio aditivo?

3. ¿Cuál es la diferencia esencial entre los dos principios?

4. ¿En dónde es menor el número de formas de ordenar objetos, en permutaciones lineales o en permutaciones circulares? Justificar la respuesta.

5. Explica la diferencia entre el resultado obtenido de utilizar permutaciones (nPr) y combinaciones (nCr).

6. Una estudiante tiene 4 blusas, 5 faldas y 2 pares de zapatos. ¿Cuántas combinaciones diferentes se pueden realizar?

7. Hay dos caminos de A y B, tres de C a B y cinco de A a C. ¿Cuántas formas hay de salir de B a una ciudad inmediata?

8. ¿Cuántos números diferentes que contengan los dígitos 3, 8, 1 y 2 se pueden formar si deben ser de cuatro cifras?

9. ¿De cuántas formas es posible ordenar 6 libros en un lugar donde sólo hay espacio para tres?

10. ¿Cuántas maneras diferentes hay de ordenar las letras de la palabra *propositivo*?

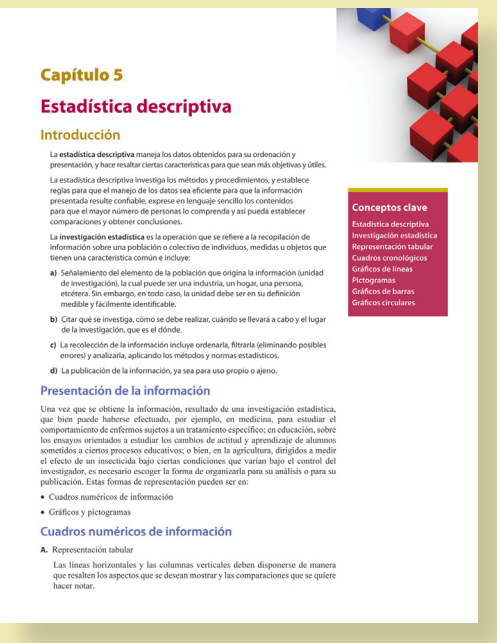
Ejercicios de repaso

Con esta sección de ejercicios concluyes el estudio de un capítulo. Los problemas a realizar en este apartado incluyen aplicaciones de todos los temas analizados. Sirve como una herramienta de autoevaluación y guía de estudio.

Diagramas, gráficos y pictogramas

Para reforzar el capítulo de estadística descriptiva, mejoramos la presentación de los pictogramas, gráficas y esquemas de organización de datos.

Con estos recursos te será más fácil entender la forma de organizar la información para su análisis o publicación.



Capítulo 5

Estadística descriptiva

Introducción

La **estadística descriptiva** maneja los datos obtenidos para su ordenación y presentación, y hace resaltar ciertas características para que sean más objetivas y útiles.

La estadística descriptiva investiga los métodos y procedimientos, y establece reglas para que el manejo de los datos sea eficiente para que la información presentada resulte confiable, expresa en lenguaje sencillo los contenidos para que el mayor número de personas lo comprenda y así pueda establecer comparaciones y obtener conclusiones.

La **investigación estadística** es la operación que se refiere a la recopilación de información sobre una población o colectivo de individuos, medidas u objetos que tienen una característica común e incluye:

a) Señalamiento del elemento de la población que origina la información (unidad de investigación), la cual puede ser una industria, un hogar, una persona, etcétera. Sin embargo, en todo caso, la unidad debe ser en su definición medible y fácilmente identificable.

b) Citar qué se investiga, cómo se debe realizar, cuándo se llevará a cabo y el lugar de la investigación, que es el dónde.

c) La recolección de la información incluye ordenarla, filtrarla (eliminando posibles errores) y analizarla, aplicando los métodos y normas estadísticos.

d) La publicación de la información, ya sea para uso propio o ajeno.

Conceptos clave
Estadística descriptiva
Investigación estadística
Representación tabular
Cuadros cronológicos
Gráficos de líneas
Pictogramas
Gráficos de barras
Gráficos circulares

Presentación de la información

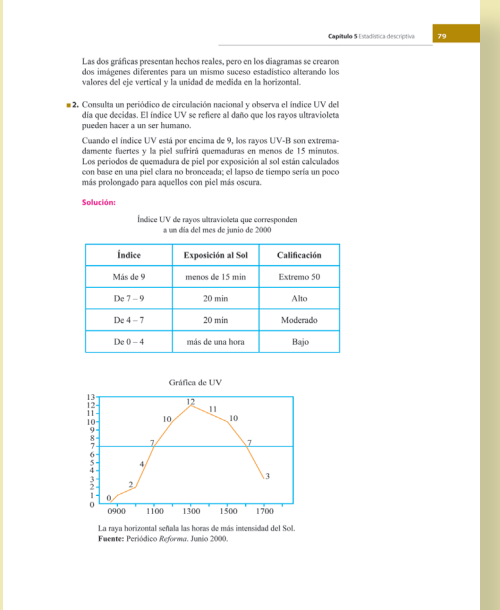
Una vez que se obtiene la información, resultado de una investigación estadística, que bien puede haberse efectuado, por ejemplo, en medicina, para estudiar el comportamiento de enfermos sujetos a un tratamiento específico; en educación, sobre los ensayos orientados a estudiar los cambios de actitud y aprendizaje de alumnos sometidos a ciertos procesos educativos; o bien, en la agricultura, dirigidos a medir el efecto de un insecticida bajo ciertas condiciones que varían bajo el control del investigador, es necesario escoger la forma de organizarla para su análisis o para su publicación. Estas formas de representación pueden ser en:

- Cuadros numéricos de información
- Gráficos y pictogramas

Cuadros numéricos de información

A. Representación tabular

Las líneas horizontales y las columnas verticales deben disponerse de manera que resalten los aspectos que se desean mostrar y las comparaciones que se quiere hacer notar.



Capítulo 5 Estadística descriptiva 79

Las dos gráficas presentan hechos reales, pero en los diagramas se crearon dos imágenes diferentes para un mismo suceso estadístico alterando los valores del eje vertical y la unidad de medida en la horizontal.

2. Consulta un periódico de circulación nacional y observa el índice UV del día que decidas. El índice UV se refiere al daño que los rayos ultravioleta pueden hacer a un ser humano.

Cuando el índice UV está por encima de 9, los rayos UV-B son extremadamente fuertes y la piel sufrirá quemaduras en menos de 15 minutos. Los períodos de quemadura de piel por exposición al sol están calculados con base en una piel clara no bronceada; el lapso de tiempo sería un poco más prolongado para aquellos con piel más oscura.

Solución:

Índice UV de rayos ultravioleta que corresponden a un día del mes de junio de 2000

Índice	Exposición al Sol	Calificación
Más de 9	menos de 15 min	Extremo 50
De 7 – 9	20 min	Alto
De 4 – 7	20 min	Moderado
De 0 – 4	más de una hora	Bajo

Gráfica de UV

La raya horizontal señala las horas de más intensidad del Sol.
Fuente: Periódico *Reforma*. Junio 2000.

Contenido

Capítulo 1 Conjuntos 1

Introducción 1
Determinación de un conjunto 1
Relación de pertenencia 2
Conjunto vacío 2
Conjunto universal 2
Conjunto de conjuntos 3
Conjunto potencia (número de subconjuntos de un conjunto) 3
Relación de conjuntos 4
Conjuntos iguales 4
Desigualdad de conjuntos 4
Conjuntos finitos o infinitos 4
Operaciones entre conjuntos 5
Unión 5
Intersección 6
Conjuntos disjuntos 6
Diferencia entre conjuntos 8
Complemento de un conjunto 8
Conjunto producto 9
Diagrama de árbol 10
Diagramas de Venn-Euler 11
Ejercicios de repaso 17

Capítulo 2 Leyes de las operaciones con conjuntos y sus aplicaciones 19

Introducción 19
Leyes de la idempotencia 19
Leyes asociativas 19
Leyes conmutativas 20
Leyes distributivas 21
Leyes de identidad (unión e intersección de conjuntos) 21
Leyes de complemento 22
Leyes De Morgan 24
Problemas resueltos 26
Ejercicios de repaso 35

Capítulo 3 Análisis combinatorio 37

Introducción 37
Principios fundamentales del conteo 37

Principio multiplicativo 37
Principio aditivo 40
Factorial 43
Permutaciones 44
Permutaciones lineales 44
Permutaciones de *n* elementos, no todos diferentes entre sí 46
Permutaciones circulares (cíclicas) 47
Combinaciones 49
Relaciones de las permutaciones y las combinaciones 51
Resumen 54
Problemas resueltos 55
Ejercicios de repaso **64**

Capítulo 4 Teorema del binomio. Triángulo de Tartaglia. Triángulo de Pascal **65**

Teorema del binomio 65
Triángulo de Tartaglia 66
Triángulo de Pascal 67
Problemas resueltos 67
Ejercicios de repaso **71**

Capítulo 5 Estadística descriptiva **73**

Introducción 73
Presentación de la información 73
Cuadros numéricos de información 73
Gráficos y pictogramas 77
Gráficos de barras 82
Gráficos circulares 83
Ejercicios de repaso **85**

Capítulo 6 Probabilidad **87**

Introducción 87
Probabilidad como frecuencia relativa 87
Consideraciones generales 87
Probabilidad expresada en tanto por ciento 88
Propiedades de la frecuencia relativa 89
Probabilidad de que ocurra o no un suceso 89
Datos de un problema 90
Población 90
Experimento aleatorio 91
Muestra 91

Tipos de sucesos 93
Probabilidad con base en los sucesos compuestos. Probabilidad axiomática 96
Consideraciones generales 96
Unión de conjuntos 96
Intersección de conjuntos 96
Diferencia de sucesos 96
Ley multiplicativa de la probabilidad 101
Uso de las leyes aditivas y multiplicativas de la probabilidad 102
Probabilidad de una diferencia 109
Ventaja de un suceso 112
Resumen 114
Probabilidad como frecuencia relativa 114
Probabilidad con base en sucesos compuestos. Probabilidad axiomática 115
Probabilidad condicional 116
Consideraciones generales 116
Propiedades 118
Problemas resueltos 126
Resumen 135

Capítulo 7 Análisis combinatorio y probabilidad. Procesos estocásticos. Regla de Bayes 137

Análisis combinatorio y probabilidad 137
Procesos estocásticos 145
Regla de Bayes 149
Razonamiento para obtener la regla de Bayes 149

Capítulo 8 Estadística inferencial 157

Conceptos básicos 157
Población y muestra 157
Métodos estadísticos 158
Concepto de variable 158
Notación 159
Variables discretas o continuas 159
Organización de datos 160
Distribuciones del tipo uno 160
Distribuciones del tipo dos 160
Distribuciones del tipo tres 161
Marca de clase 163
Gráficas 164
Diagrama de frecuencias de puntos 165
Diagrama de barras 165

Histogramas. Datos agrupados 166
Polígonos de frecuencias 167
Curvas de frecuencia 168
Frecuencias acumuladas. Ojivas 170
Distribuciones de frecuencias relativas 171
Distribuciones porcentuales acumuladas 173
Percentiles y rango percentil 174
Ejercicios de repaso **175**

Capítulo 9 Medidas de tendencia central **177**

Generalidades 177
Parámetro 177
Media aritmética 177
Media aritmética de una distribución de frecuencias agrupadas 179
Mediana y moda 181
  Mediana 181
  Moda 184
  Moda de datos agrupados 184
Uso de la media, la mediana y la moda 185
Media geométrica y media armónica 187
Ejercicios de repaso **192**

Capítulo 10 Medidas de dispersión **195**

Generalidades 195
Rango 196
Cuartiles y deciles 197
Rango intercuartil 199
Desviación media y varianza 199
Ejercicios de repaso **206**

Capítulo 11 Desviación estándar o típica **207**

Definición 207
Dispersión relativa. Coeficientes de variación 210
Ejercicios de repaso **211**

Capítulo 12 Distribución de probabilidades discretas. Binomial o de Bernoulli. De Poisson **213**

Binomial 213
Distribución de Poisson 216
Ejercicios de repaso **221**

Capítulo 13 Distribución de probabilidades continuas. Variable normalizada. Distribución normal 223

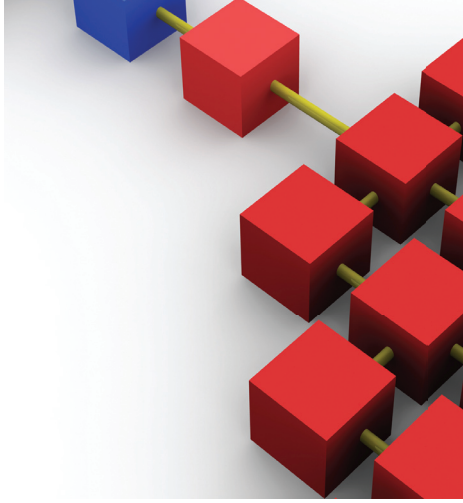
Variable normalizada. Calificación estándar Z. 223
Propiedades de la calificación estándar 224
Distribución normal 226
Propiedades de la curva normal 227
Tabla de áreas bajo la curva normal. Cómo usarla 228
Área bajo la curva 228
Cálculo del valor o valores de Z 231
Cálculo del rango percentil 234
Ejercicios de repaso 238

Capítulo 14 Correlación y regresión 239

Repaso de geometría analítica. 239
La línea recta 239
Correlación 241
Coeficientes de correlación 241
Coeficiente r de correlación lineal del producto momento (Pearson) 243
Coeficiente de correlación r por rangos de Spearman 243
Regresión 246
Ajuste de curvas. Método de mínimos cuadrados 248
Recta de regresión de mínimos cuadrados 249
Ejercicios de repaso 252

Capítulo 15 Inferencia estadística. Conceptos básicos 253

Generalidades 253
Muestreo 253
Procedimientos de muestreo 254
Muestreo aleatorio con y sin reemplazo 255
Muestreo por conglomerados 255
Muestreo estratificado 256
Muestreo sistemático 257
Distribución de las medias de las muestras 258
Estimación. Puntual y por intervalos 260
Comprobación de hipótesis (prueba de hipótesis) 262
Errores de tipo I y de tipo II 263
Ejercicios de repaso 263



Capítulo 1

Conjuntos

Introducción

La **teoría de conjuntos** es un instrumento matemático útil para la sistematización de nuestra forma de pensar porque permite la capacidad de análisis y comprensión de las interrelaciones que existen entre todas las partes de un problema y así facilitar su solución.

Analizar el tema de conjuntos en el curso de aritmética y álgebra nos permitió desarrollar los temas de operaciones con números reales y el de relaciones y funciones. En este curso daremos un repaso a esos conceptos y ampliaremos algunos aspectos para facilitar el estudio de la probabilidad y la estadística.

Aceptamos como nociones intuitivas y, por consiguiente, no definibles las de unidad, conjunto, pertenencia a un conjunto, correspondencia y orden.

Las ideas de unidad y pluralidad (conjunto) las adquiere cada ser humano en los comienzos de su vida cuando se manifiesta una de sus facultades: la diferenciación.

Los conceptos primarios de unidad y de conjunto son correlativos, es decir, no pueden concebirse por separado. Lo mismo sucede con las nociones, tales como alto y bajo, cerca y lejos, grande y pequeño.

Un **conjunto** es cualquier colección de objetos bien definidos, de tal manera que se pueda decir siempre si un objeto pertenece o no al conjunto al cual nos referimos.

Conceptos clave

- Teoría de conjuntos
- Conjunto
- Conjunto vacío
- Conjunto universal
- Conjunto de conjuntos
- Conjunto potencia
- Subconjuntos propios
- Conjuntos iguales
- Desigualdad de conjuntos
- Conjuntos finitos e infinitos
- Unión
- Intersección
- Conjuntos disjuntos
- Diferencia entre conjuntos
- Complemento de un conjunto
- Conjunto producto
- Diagrama de árbol
- Diagramas de Venn-Euler

Determinación de un conjunto

Los conjuntos se denotan con letras mayúsculas y es posible determinar o establecer un conjunto por *enumeración* o *descripción*.

- **Enumeración** (también se le llama *extensión*). En este método los elementos que lo integran se colocan dentro de este tipo de llaves { } y separados por comas. Por ejemplo:

$$A = \{3, 4, 5\}$$

$$B = \{\text{Luis, Pedro, Ignacio}\}$$

- **Descripción** (también se le llama *comprensión*). En esta forma se enuncia una propiedad o atributo que caracterice a todos los elementos del conjunto. Por ejemplo:

$$D = \{\text{los números enteros menores que } -2\}$$

$$F = \{\text{los divisores del } 21\}$$

Otra forma más práctica de definir conjuntos, también por descripción, es aquella que consiste en el uso de una variable genérica, por ejemplo x; es decir, un indicador de elementos y una frase o relación matemática que especifique con toda precisión los elementos que se estén generando, todo ello encerrado en llaves.

Además, se usa el símbolo " | ", que dentro de la teoría de conjuntos se lee "tal que".

Ejemplos:

- **1.** $A = \{x \mid x \text{ es una vocal}\}$, de donde $A = \{a, e, i, o, u\}$
- **2.** $H = \{x \mid x + 7 = 10\}$, de donde, y resolviendo la ecuación $H = \{3\}$
- **3.** $J = \{x \mid x^2 + 6x + 8 = 0\}$ de donde, y resolviendo la ecuación $J = \{2, 4\}$

Relación de pertenencia

Dado el conjunto $A = \{1, 2, 3\}$ para expresar que el número 2 es un elemento del conjunto A se emplea el símbolo $\in$, el cual se lee "es un elemento de" o "pertenece a"; por lo tanto, se indica:

$$2 \in A$$

Si queremos expresar que los números 1 y 3 son elementos del conjunto A queda:

$$1, 3 \in A \text{ o también } 1 \in A, 3 \in A.$$

Cuando un elemento no pertenece a un conjunto se usa el símbolo $\notin$, que se lee "no es elemento de" o "no pertenece a". Por ejemplo, sea el siguiente conjunto:

$J = \{x \mid 10 < x \leq 15, x \in N\}$, de donde $J = \{11, 12, 13, 14, 15\}$

La letra N identifica a los números naturales.

Para indicar que el número 8 no pertenece al conjunto J se escribe:

$$8 \notin J$$

Conjunto vacío

Los conjuntos que no tienen elementos se denominan conjuntos vacíos y se representan con el símbolo $\varnothing$. Por ejemplo, sea H el conjunto de los números naturales pares mayores que 2 y menores que 4.

$$H = \{x \mid 2 < x < 4, x \in N \text{ par}\}, \text{ de donde } H = \varnothing$$

No debe expresarse como $H = \{\varnothing\}$

El conjunto vacío también se puede expresar con las llaves vacías:

$$H = \{\ \}$$

Conjunto universal

Si $U \neq \varnothing$ es cierto conjunto cuyos subconjuntos están en consideración, se dice que el conjunto dado es un **conjunto universal**. El símbolo con el que se representa es U.

Ejemplo:

- **1.** Sea el conjunto $U = \{\text{los estados de la República Mexicana}\}$, los subconjuntos serían, entre otros, los siguientes:

  $A = \{\text{Tlaxcala, Aguascalientes}\}$

  $B = \{\text{Durango}\}$

En ocasiones se citan los conjuntos sin ninguna otra indicación y sin saber a qué conjunto U pertenecen.

Ejemplo:

1. $C = \{2, 3, 4, 5\}$

Entre otros, el conjunto U podría ser:

$U = \{1, 2, 3, \ldots, 10\}$

$N = \{\text{los números naturales}\}$

Conjunto de conjuntos

Los elementos de un conjunto son, a su vez, conjuntos, lo que hace pensar en **conjunto de conjuntos**. Por ejemplo, un año es un conjunto de conjuntos porque el año es un conjunto de meses y éstos, a su vez, lo son de semanas y éstas, de días.

Conjunto potencia (número de subconjuntos de un conjunto)

A todos los subconjuntos de un conjunto se les llama **conjunto potencia** y se expresa $P(A)$.

Ejemplo:

1. Dado el conjunto $A = \{a, b, c\}$, determina cuáles subconjuntos se pueden formar.

Solución:

$P(A) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$

El conjunto $\{a, b, c\}$ es un subconjunto de A porque todos sus elementos pertenecen a dicho conjunto, es decir:

$$A \subseteq A$$

Además, el conjunto vacío es un subconjunto de cualquier conjunto:

$$\varnothing \subset A$$

La cardinalidad n del conjunto potencia $P(A)$ se obtiene con 2^n, donde n es el número de elementos del conjunto A y se denota con $n[\,P(A)]$

Continuamos con el conjunto citado en este subpárrafo:

$$A = \{a, b, c\}$$

Su cardinalidad es:

$$n\,(A) = 3$$

La cardinalidad del conjunto potencia $P(A)$ es:

$n[\,P(A)] = 2^n = 2^3 = 8$, que es el mismo resultado que obtuvimos.

Los subconjuntos de un conjunto (sin considerar el conjunto que lo genera) se llaman **subconjuntos propios** y hay tantos como $2^n - 1$, donde n también es el número de elementos del conjunto.

Así, los subconjuntos propios del conjunto $A = \{a, b, c\}$ son:

$$2^n - 1 = 2^3 - 1 = 8 - 1 = 7$$

Si necesitamos citar cuáles son, tenemos:

$$\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}$$

Relación de conjuntos

Conjuntos iguales

Para que dos conjuntos sean iguales deben tener los mismos elementos y en consecuencia, debe cumplirse simultáneamente:

$$A \subseteq B \text{ y } B \subseteq A$$

Esta relación se indica con el símbolo "=" y se lee "igual a" o "es igual a"

Ejemplo:

- **1.** $A = B \Leftrightarrow A \subseteq B$ y $B \subseteq A$

Desigualdad de conjuntos

Sean los conjuntos:

$$A = \{1, 2, 3\}$$

$$B = \{1, 2, 3, 4\}$$

No se cumple en forma simultánea $A \subseteq B$ y $B \subseteq A$

Esta relación se indica con el símbolo de desigualdad "≠", que se lee, "es desigual a" o "es diferente".

Ejemplo:

- **1.** $A \neq B$

Conjuntos finitos e infinitos

Un conjunto es finito cuando sus elementos se pueden poner en correspondencia biunívoca con un subconjunto de los primeros K números naturales; si no es así, el conjunto es infinito.

Ejemplos:

- **1.** $M = \{x \mid 6 < x < 75, x \in N\}$, de donde:

  $M = \{7, 8, 9, \ldots, 74\}$ es un conjunto finito con 68 elementos

- **2.** $H = \{N\}$, de donde:

  $H = \{1, 2, 3, \ldots,\}$ es un conjunto infinito

- **3.** $J =$ {los múltiplos de 5} de donde:

  $J = \{5, 10, 15, \ldots,\}$ conjunto infinito

Operaciones entre conjuntos

Las operaciones con conjuntos son formas específicas de combinarlos para obtener otros conjuntos. Todas las operaciones entre conjuntos son binarias.

Las operaciones son:

A. Unión

B. Intersección

C. Conjuntos disjuntos

D. Diferencia entre conjuntos

E. Complemento de un conjunto

F. Conjunto producto

G. Diagrama de árbol

Unión

Si se reúnen los elementos de dos o más conjuntos para formar uno solo, a este conjunto se le denomina *unión de conjuntos*. Si existen elementos comunes entre los conjuntos originales, éstos no se repiten en el conjunto unión.

La unión se representa con el símbolo $\cup$ colocado entre los conjuntos. Así, $A \cup B$ se lee "unión de A y B" o "A unión B".

Cuando el conjunto se establece por descripción usando el símbolo "tal que", la unión se expresa de la siguiente forma:

$$A \cup B = \{x \mid x \in A \text{ o } x \in B\}$$

El conectivo lógico "o" que relaciona a las dos condiciones es una *o* inclusiva.

Ejemplo:

1. Sean los conjuntos:

$P = \{1, 2, 3, 4\}$

$M = \{3, 4, 5, 6\}$

$P \cup M = \{1, 2, 3, 4, 5, 6\}$

Propiedades de la unión de conjuntos

$A \cup B = B \cup A$

$A \cup (B \cup C) = (A \cup B) \cup C$

$A \cup \varnothing = A$

$A \cup U = U$

$A \cup A = A$

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

$A \cup (A \cap B) = A$

Si $A \cup B = \varnothing$ entonces $A = \varnothing$ y $B = \varnothing$

A y B son ambos subconjuntos de $A \cup B$, esto es:

$A \subset (A \cup B)$ y $B \subset (A \cup B)$

Intersección

La intersección de dos conjuntos A y B es el conjunto que forman los elementos comunes a ambos conjuntos. Se representa con el símbolo $\cap$ colocado entre los conjuntos. Así, $A \cap B$ se lee "intersección de A y B" o "A intersección B".

Cuando el conjunto se determina por descripción usando el símbolo "tal que" la intersección se expresa en la forma siguiente:

$A \cap B = \{x \mid x \in A \text{ y } x \in B\}$ o también:

$A \cap B = \{x \mid x \in A,\ x \in B\}$, donde la coma tiene el significado de *y* copulativa

Ejemplos:

1. Sean los conjuntos:

$A = \{1, 2, 3, 4\}$

$B = \{1, 2, 5, 6\}$

$A \cap B = \{1, 2\}$

2. Sean los conjuntos:

$P = \{1, 2, 3\}$

$M = \{6, 7\}$

$P \cap M = \varnothing$

Propiedades de la intersección de conjuntos

$A \cap B = B \cap A$

$A \cap \varnothing = \varnothing$

$A \cap U = A$

$A \cap A = A$

$(A \cap B) \cap C = A \cap (B \cap C)$

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$A \cap (A \cup B) = A$

Cada uno de los conjuntos A y B contienen a $A \cap B$ como subconjuntos, es decir:

$$(A \cap B) \subset A \text{ y } (A \cap B) \subset B$$

Conjuntos disjuntos

Si dos conjuntos A y B no tienen elementos comunes, es decir, si ningún elemento de A está en B y si ningún elemento de B está en A, entonces A y B son disjuntos o ajenos entre sí y su intersección es el conjunto vacío:

$$A \cap B = \varnothing$$

Ejemplos:

■ **1.** Con los conjuntos:

$A = \{1, 2, 3\}$

$B = \{0, 4, 5\}$

Los conjuntos A y B son disjuntos: $A \cap B = \varnothing$

■ **2.** Con los conjuntos:

$M = \{\text{los números enteros positivos}\}$

$N = \{\text{los números enteros negativos}\}$

Los conjuntos M y N son disjuntos: $M \cap N = \varnothing$

■ **3.** Con los conjuntos:

$C = \{1, 2, 4\}$

$D = \{0, 4\}$

Los conjuntos C y D no son disjuntos porque $4 \in C$ y $4 \in D$ y $C \cap D = \{4\} \neq \varnothing$

Uso de paréntesis

Los paréntesis indican qué operación se debe hacer primero. En general, procedemos en forma semejante a como lo explicamos en los cursos anteriores: "Cuando una expresión algebraica contiene uno o más pares de símbolos de agrupación, encerrados en otro par, siempre se elimina primero el de más adentro".

Ejemplos:

■ **1.** Sean los conjuntos:

$T = \{1, 2, 3\}$

$P = \{1, 3, 4, 5\}$

$L = \{5, 6, 7\}$

Obtener:

$(T \cup P) \cap L =$

Inicialmente obtenemos $T \cup P$

$T \cup P = \{1, 2, 3, 4, 5\}$

Ahora debemos obtener la intersección con el conjunto L:

$(T \cup P) \cap L = \{5\}$

■ **2.** Usando los mismos conjuntos señalados, determina:

$T \cup P\ (P \cap L)$

Inicialmente obtenemos:

$P \cap L = \{5\}$

Ahora debemos realizar la unión con el conjunto T:

$T \cup (P \cap L) = \{1, 2, 3, 5\}$

Nota: La operación $(T \cup P) \cap L$ es distinta de $T \cup (P \cap L)$.

Diferencia entre conjuntos

Dados los conjuntos A y B, el conjunto *diferencia* se define como la diferencia de $A - B$; en este orden, es el conjunto de todos los elementos que *pertenecen* a A pero no a B. La diferencia de A y B se expresa de la siguiente forma:

$A - B$, que se lee "A diferencia B" o "A menos B".

Cuando el conjunto se determina por descripción usando el símbolo "tal que", la diferencia se expresa así:

$$A - B = \{ x \in U \mid x \in A, x \notin B\}$$

o también:

$$A - B = \{ x \mid x \in A, x \notin B\}$$

Algunos autores expresan la diferencia de A y B con:

$$A / B \text{ o } A \sim B$$

Ejemplo:

1. $A = \{1, 2, 3, 4, 5\}$

$B = \{1, 2\}$

$A - B = \{3, 4, 5\}$

Observa las operaciones siguientes en que aplicamos la diferencia entre conjuntos:

$(A - B) \subset A$, el conjunto A contiene al $A - B$ como subconjunto.

Los conjuntos $(A - B)$, $(B - A)$ y $A \cap B$ son mutuamente disjuntos, es decir, la intersección de dos cualesquiera de ellos es el conjunto vacío.

Complemento de un conjunto

Cuando se ha establecido un conjunto universal U, a la diferencia de U y a la de un conjunto (sea por ejemplo A) se le llama complemento de A y se expresa A'. El apóstrofe señala que hemos formado el complemento de A.

Algunos autores expresan el complemento así:

$$A^c$$

de donde $A' = A_c$. Cuando el conjunto complemento se cita por descripción usando el símbolo "tal que" queda:

$$A' = \{ x \in U \mid x \notin A \}$$

o también:

$$A' = \{ x \mid x \notin A \}$$

Ejemplo:

1. $U = \{1, 2, 3, 4, 5, 6\}$

$A = \{1, 2, 3\}$

$A' = \{4, 5, 6\}$

Observa las siguientes operaciones en que aplicamos el complemento de un conjunto:

$A \cap A' = \varnothing$

$A \cup A' = U$

$U' = \varnothing$

$\varnothing' = U$

$(A')' = A$ el complemento del complemento de un conjunto A es el conjunto A.

Para todo conjunto $A \subset U$ se tiene que el complemento $A' = U - A$. En la operación con conjuntos $A - B$, su resultado es la resta de A y B.

Conjunto producto

En tu curso de *aritmética y álgebra* se estableció, en el tema de conjunto producto (producto cartesiano) que:

Sean los conjuntos:

$A = \{a, e\}$

$B = \{1, 2, 3\}$

El producto cartesiano de estos dos conjuntos $A \times B$, en este orden, es el conjunto de todos los posibles pares ordenados, tales que la primera componente del par ordenado es un elemento de A y la segunda componente es un elemento de B.

La expresión $A \times B$ se lee "A cruz B" y se expresa, por descripción, así:

$$A \times B = \{ (x, y) \mid x \in A, y \in B \}$$

Esta expresión se lee:

La pareja (x, y), tal que x pertenece al conjunto A y y pertenece al conjunto B.

Si se desarrolla el producto de los conjuntos citados obtenemos:

$$A \times B = \{(a, 1), (a, 2), (a, 3), (e, 1), (e, 2), (e, 3)\}$$

Los elementos del conjunto producto son parejas ordenadas:

$$\{(a, 1), (a, 2), (a, 3), (e, 1), (e, 2), (e, 3)\}$$

En la pareja $(a, 1)$, a se denomina *primera componente* y el número 1 se conoce como *segunda componente*.

En el caso en que los elementos de los conjuntos sean números reales, es costumbre llamar a la primera componente de la pareja ordenada *abscisa* y a la segunda *ordenada*. Con estos conceptos iniciaste tu curso de *geometría analítica*.

Diagrama de árbol

Si en un problema es necesario obtener el producto de tres o más conjuntos, el desarrollo resulta complicado, se utiliza el **diagrama de árbol.**

Los diagramas de árbol se trazan horizontalmente de la manera siguiente:

1. Se inicia el diagrama con tantas ramificaciones primarias como elementos tenga el primer conjunto y se anota en cada extremo terminal primario uno de los elementos del primer conjunto.

2. En cada extremo terminal primario se trazan tantas ramificaciones como elementos tenga el segundo conjunto y se anotan en cada rama los elementos del segundo conjunto, y así sucesivamente hasta incluir todos los conjuntos que intervienen en la operación.

3. Finalmente, para obtener los agrupamientos del resultado se recorre el diagrama desde su inicio hasta todas y cada una de las terminales finales, agrupando como uno solo los elementos simples que se encuentran en cada recorrido.

Ejemplo:

1. Sean los conjuntos:

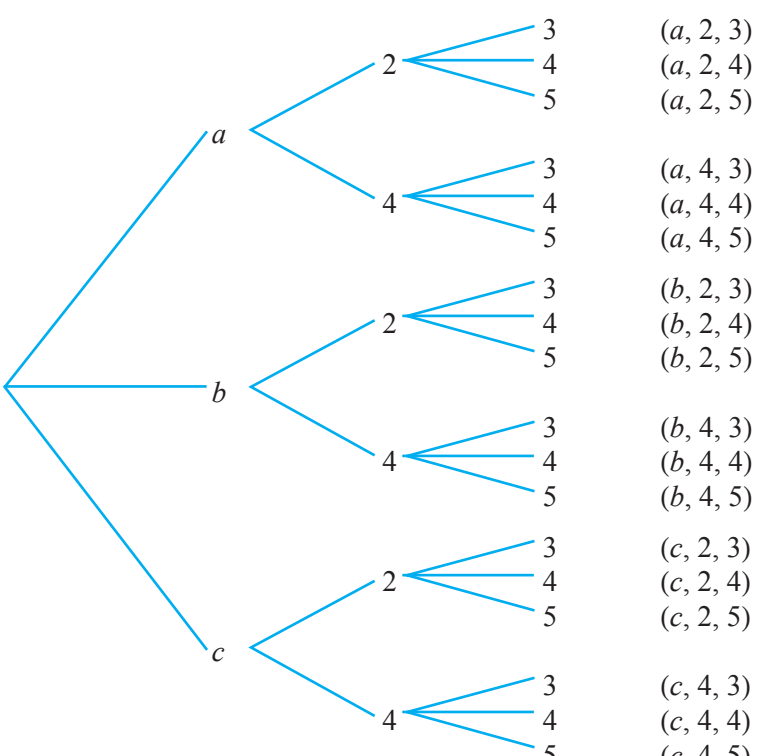
$A = \{a, b, c\}$

$B = \{2, 4\}$

$C = \{3, 4, 5\}$

Determina el conjunto producto $A \times B \times C$ con el diagrama de árbol.

El resultado de este producto es el conjunto de las ternas y se listan a la derecha.



Solución:

$A \times B \times C = \{(a, 2, 3), (a, 2, 4)\ (a, 2, 5), (a, 4, 3), (a, 4, 4), (a, 4, 5), (b, 2, 3), (b, 2, 4), (b, 2, 5), (b, 4, 3), (b, 4, 4), (b, 4, 5), (c, 2, 3), (c, 2, 4), (c, 2, 5), (c, 4, 3), (c, 4, 4), (c, 4, 5)\}$.

Si el conjunto A tiene n elementos, el conjunto B tiene m elementos y el conjunto C tiene q elementos, el conjunto producto $A \times B \times C$ tendrá nmq elementos.

Si uno de los conjuntos A, B o C es un conjunto vacío, el resultado de $A \times B \times C$ será un conjunto vacío.

En el ejemplo anterior, el número de elementos de A, B y C es de 3, 2 y 3 elementos, respectivamente. Así, el conjunto producto tiene $3(2)(3) = 18$ elementos, misma cantidad de elementos que obtuvimos en el resultado.

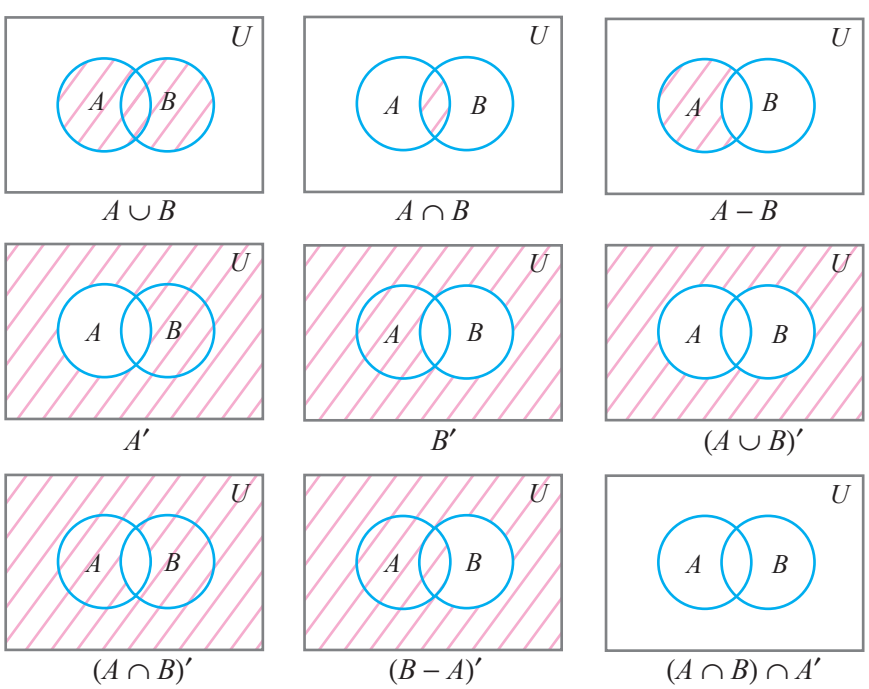
Diagramas de Venn-Euler

Son representaciones gráficas de los conjuntos que nos permiten visualizarlos mejor.

El conjunto universal U está representado por puntos que, por cierto, no se indican en el interior de un rectángulo.

Ejemplos:

1. En las siguientes operaciones, el área sombreada es el resultado de cada una, excepto en el último porque el resultado es el conjunto vacío.

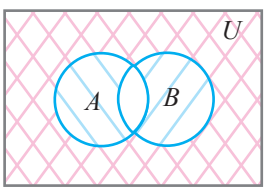


2. Determina $A' \cap B'$:

Solución:

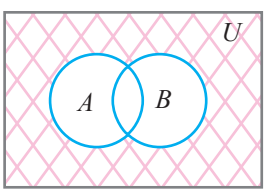
Primero obtenemos A', que es la parte exterior de A, con trazos inclinados de derecha a izquierda.

A continuación determinamos B', que es la parte exterior de B con trazos inclinados de izquierda a derecha.



Solución:

$A' \cap B'$es el área con doble rayado.

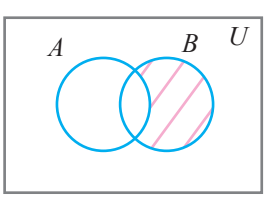


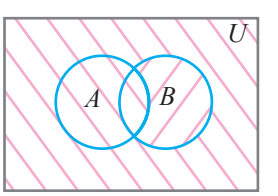
3. Determina $(B - A)'$:

Solución:

Primero obtenemos $B - A$ y marcamos con trazos inclinados de derecha a izquierda lo que está en B pero no está en A.

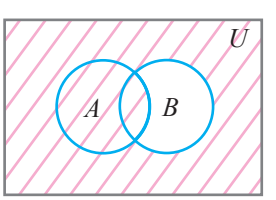
A continuación determinamos $(B - A)'$ y marcamos con trazos inclinados de izquierda a derecha.





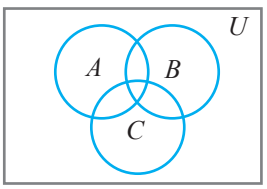
Solución:

La parte rayada del conjunto U es $(B - A)'$:



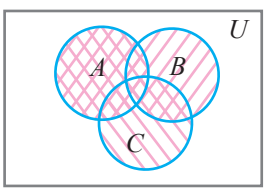
En los ejercicios siguientes, en cada uno de los diagramas de Venn A, B, C que se incluyen, expresa el resultado de las operaciones que se citan.

■ **4.** Determina $(A \cup B) \cap (A \cup C)$ en:



Solución:

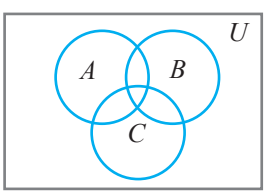
Sombreamos $A \cup B$ con trazos inclinados de derecha a izquierda. A continuación, $A \cup C$ con trazos inclinados de izquierda a derecha.



Solución:

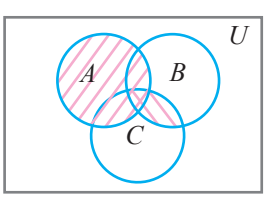
La parte de los conjuntos que tienen doble raya es $(A \cup B) \cap (A \cup C)$.

■ **5.** Determina $A \cup (B \cap C)$ en:



Solución:

Sombreamos A con trazos inclinados de derecha a izquierda; en seguida $B \cap C$ con trazos inclinados de izquierda a derecha.

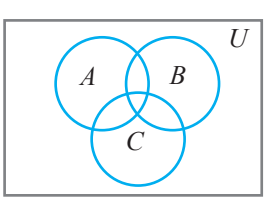


Solución:

La parte rayada de los conjuntos es $A \cup (B \cap C)$.

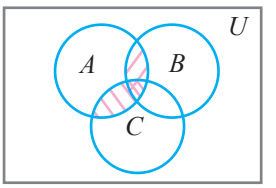
Observa: $(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$, ejemplos 4 y 5.

■ **6.** Determina $(A \cap B) \cup (A \cap C)$ en:



Solución:

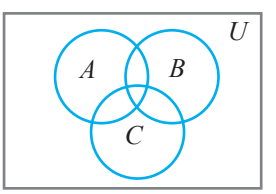
Marcamos $A \cap B$ con trazos inclinados de derecha a izquierda. En seguida, marcamos $A \cap C$ con trazos inclinados de izquierda a derecha.



Solución:

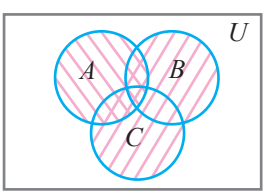
La parte rayada de los conjuntos es $(A \cap B) \cup (A \cap C)$.

■ **7.** Determina $A \cap (B \cup C)$ en:



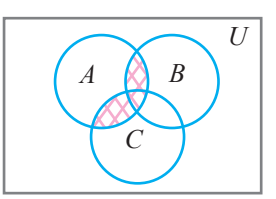
Solución:

Marcamos $B \cup C$ con trazos inclinados de derecha a izquierda. A continuación marcamos A con trazos inclinados de izquierda a derecha.



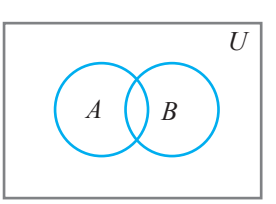
Solución:

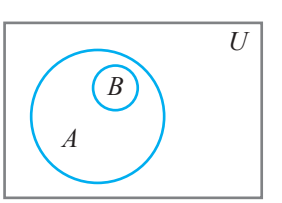
La parte de los conjuntos que tienen raya doble es $A \cap (B \cup C)$.

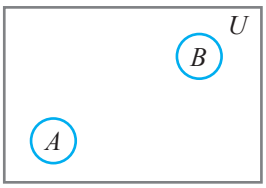


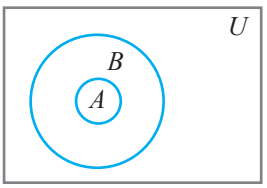
Observa: $(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$, ejemplos 6 y 7.

■ **8.** Sombrea $A \cup B$ en los siguientes diagramas:



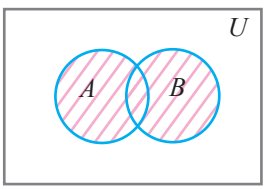


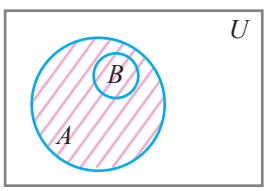


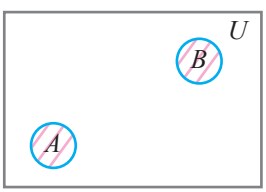


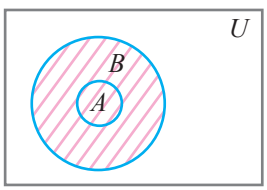
Solución:

La parte de los conjuntos rayados es $A \cup B$:

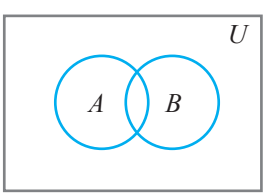


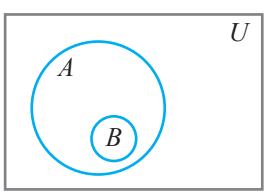


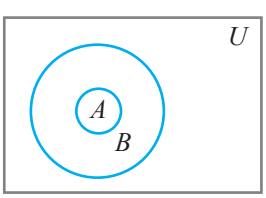


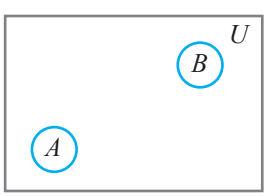


9. Sombrea $A \cap B$ en los siguientes diagramas:



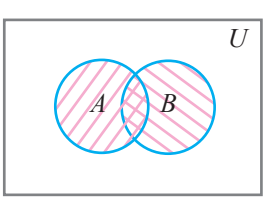


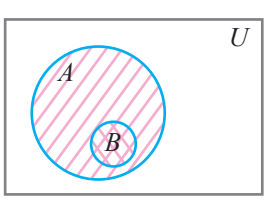


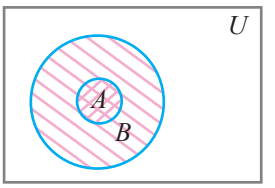


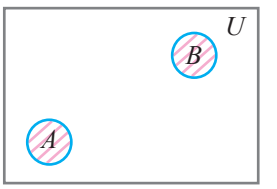
Solución:

Como la intersección de A y B es el área común de A y de B, para obtener $A \cap B$ marcamos primero A con trazos inclinados de derecha a izquierda y a continuación marcamos B con trazos inclinados de izquierda a derecha.



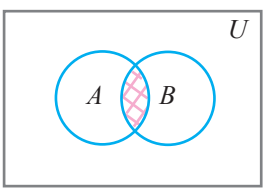


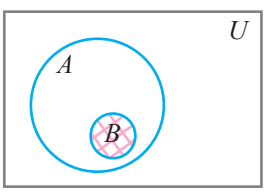


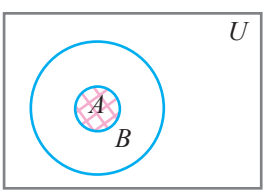


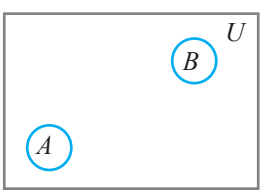
Solución:

La parte de los conjuntos que tiene doble raya es $A \cap B$.









Son disjuntos

¡Aplícate!

Con los siguientes conjuntos:

$U = \{1, 2, 3, 4, 5, 6, 7\}$

$A = \{1, 2, 3, 4, 5\}$

$B = \{1, 3, 5, 7\}$

$C = \{2, 5, 6, 7\}$

determina:

1. $A \cup C$ **Sol.** $\{1, 2, 3, 4, 5, 6, 7\}$

2. $B \cap A$ **Sol.** $\{1, 3, 5\}$

3. B' **Sol.** $\{2, 4, 6\}$

4. $C - B$ **Sol.** $\{2, 6\}$

5. $A' - B$ **Sol.** $\{6\}$

6. $B' \cup C$ **Sol.** $\{2, 4, 5, 6, 7\}$

7. $C' \cap A$ **Sol.** $\{1, 3, 4\}$

8. $(A - C)'$ **Sol.** $\{2, 5, 6, 7\}$

9. $(A \cap A')'$ **Sol.** U

10. $(A - B')'$ **Sol.** $\{2, 4, 6, 7\}$

11. Expresa el conjunto B por comprensión y por extensión si B es el conjunto de los números reales cuyos cuadrados son igual a 36.

Sol. Por comprensión $B = \{x \mid x^2 = 36\}$
Por extensión $B = \{6, -6\}$

12. Si

$U = \{1, 2, 3, 4, 5\}$
$A = \{2, 3, 5\}$
$B = \{4, 1, 5, 6\}$, verifica que:
$A - B = A \cap B'$

Sol. $A - B = \{2, 3\}$
$A \cap B' = \{2, 3\}$
$A - B = A \cap B'$

13. Con los mismos conjuntos del ejercicio anterior, verifica que:

$A = (A \cap B') \cup (A \cap B)$

Sol. $A = \{2, 3, 5\}$
$(A \cap B') \cup (A \cap B) = \{2, 3, 5\}$
$A = (A \cap B') \cup (A \cap B)$

Conclusión de los ejercicios 12 y 13:

$A = (A \cap B') \cup (A \cap B)$

$A - B = A \cap B'$

Ejercicios de repaso

Sea U el conjunto de los números racionales (Q) para los incisos 1 y 2.

1. Expresa el conjunto U por comprensión.

2. Escribe el símbolo que corresponda en cada caso ($\in$, $\notin$, $\subset$, $\not\subset$).

a) 4 _____ U **Sol.** $\in$

b) $\sqrt{2}$ _____ U **Sol.** $\notin$

c) $3i$ _____ U **Sol.** $\notin$

d) $3, -3 \_\_\_\_\_ U$ **Sol.** $\in$

e) $0 \_\_\_\_\_ U$ **Sol.** $\in$

f) $\frac{4}{5} \_\_\_\_\_ U$ **Sol.** $\in$

g) $\{\text{Números pares}\} \_\_\_\_\_ U$ **Sol.** $\subset$

h) $\{\text{Números complejos}\} \_\_\_\_\_ U$ **Sol.** $\not\subset$

i) $\{0\} \_\_\_\_\_ U$ **Sol.** $\subset$

j) $\left\{\frac{1}{2} < x < 10, x \in Q\right\} \_\_\_\_\_ U$ **Sol.** $\subset$

3. Escribe, por extensión, el conjunto A si A es el conjunto de los números dígitos.

Sol. $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

4. Escribe la cardinalidad del conjunto A.

Sol. 10

5. Escribe, por comprensión, el conjunto B si B es el conjunto de los números enteros.

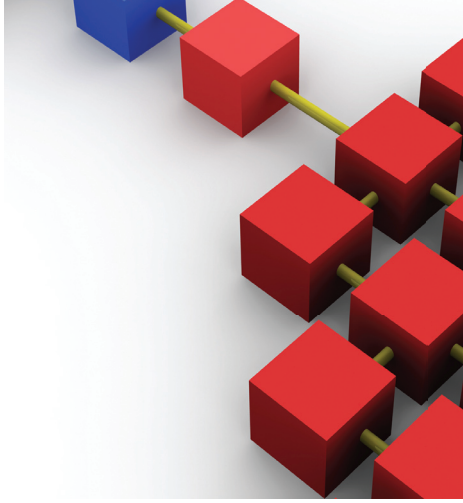
Sol. $B = \{h \times 1 \times \in z\}$

6. Escribe la cardinalidad del conjunto B.

Sol. Infinito

7. Calcula la cardinalidad del conjunto potencia $P(A)$.

Sol. 1024



Capítulo 2

Leyes de las operaciones con conjuntos y sus aplicaciones

Introducción

Las leyes que rigen las operaciones con conjuntos permiten:

a) Demostrar las operaciones

b) Simplificar una operación combinada

c) Aplicarlas en las relaciones para el cálculo de probabilidades

Conceptos clave

- Leyes de idempotencia
- Leyes asociativas
- Leyes conmutativas
- Leyes distributivas
- Leyes de identidad
- Leyes de complemento
- Leyes de De Morgan
- Ley de la diferencia de dos conjuntos

Leyes de idempotencia

a) La unión de un conjunto consigo mismo es igual al conjunto original.

$$A \cup A = A$$

Ejemplo:

■ **1.** $A = \{3, 2, 4\}$

$A \cup A = \{3, 2, 4\}$

b) La intersección de un conjunto consigo mismo es igual al conjunto original.

$$A \cap A = A$$

Ejemplo:

■ **1.** $A = \{3, 2\}$

$A \cup A = \{3, 2\}$

Leyes asociativas

Sean los conjuntos:

$A = \{3, 2\}$

$B = \{1, 3\}$

$C = \{3, 4, 5\}$

a) El resultado de la unión de dos conjuntos, unido a su vez con un tercer conjunto, es igual a la unión del primero con la unión del segundo con el tercero.

$$(A \cup B) \cup C = A \cup (B \cup C)$$

Ejemplo:

■ **1.** Sean los conjuntos citados, expresar:

$$(A \cup B) \cup C = A \cup (B \cup C)$$

Operación en el primer miembro	Operación en el segundo miembro
$(A \cup B) \cup C$	$A \cup (B \cup C)$
$A \cup B = \{1, 2, 3\}$	$B \cup C = \{1, 3, 4, 5\}$
$(A \cup B) \cup C = \{1, 2, 3, 4, 5\}$	$A \cup (B \cup C) = \{1, 2, 3, 4, 5\}$

$$\{1, 2, 3, 4, 5\} = \{1, 2, 3, 4, 5\}$$

b) Si en la intersección de dos conjuntos su resultado interseca a su vez un tercer conjunto, el resultado es igual a la intersección del primero con la intersección del segundo con el tercero.

$$(A \cap B) \cap C = A \cap (B \cap C)$$

Ejemplo:

■ **1.** Sean los conjuntos A, B y C citados, expresa:

Operación en el primer miembro	Operación en el segundo miembro
$(A \cap B) \cap C$	$A \cap (B \cap C)$
$A \cap B = \{3\}$	$B \cap C = \{3\}$
$(A \cap B) \cap C = \{3\}$	$A \cap (B \cap C) = \{3\}$

$$\{3\} = \{3\}$$

Leyes conmutativas

Sean los conjuntos:

$A = \{2, 5, 4\}$

$B = \{4, 6\}$

a) La unión de dos conjuntos es igual a la unión del segundo con el primero.

$$A \cup B = B \cup A$$

Ejemplo:

■ **1.** Sean los conjuntos A y B citados, expresa:

$A \cup B = \{2, 4, 5, 6\}$ | $B \cup C = \{2, 4, 5, 6\}$

$$\{2, 4, 5, 6\} = \{2, 4, 5, 6\}$$

b) La intersección de los conjuntos es igual a la intersección del segundo con el primero.

$$A \cap B = B \cap A$$

Ejemplo:

■ **1.** Sean los conjuntos A y B citados, expresa:

$A \cap B = \{4\}$ | $B \cap A = \{4\}$

$$\{4\} = \{4\}$$

Leyes distributivas

Sean los conjuntos:

$A = \{0, 1, 4, 6\}$

$B = \{1, 3\}$

$C = \{1, 4, 5\}$

a) En la unión de un conjunto con la intersección de otros dos conjuntos, su resultado es igual a la unión del primero con el segundo intersecada con la unión del primero con el tercero.

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Ejemplo:

■ **1.** Sean los conjuntos A, B y C citados, expresa:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$(B \cap C) = \{1\}$

$A \cup (B \cap C) = \{0, 1, 4, 6\}$

$(A \cup B) = \{0, 1, 3, 4, 6\}$

$(A \cup C) = \{0, 4, 5, 6\}$

$(A \cup B) \cap (A \cup C) = \{0, 1, 4, 6\}$

$$\{0, 1, 4, 6\} = \{0, 1, 4, 6\}$$

b) La intersección de un conjunto con la unión de otros conjuntos es igual a la intersección del primero con el segundo unida a la intersección del primero con el tercero.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Ejemplo:

■ **1.** Sean los conjuntos A, B y C citados, expresa:

$$A \cap (B \cup C) = (A \cap B\} \cup (A \cap C)$$

$(B \cup C) = \{1, 3, 4, 5\}$

$A \cap (B \cup C) = \{1, 4\}$

$(A \cap B) = \{1\}$

$(A \cap C) = \{1, 4\}$

$(A \cap B) \cup (A \cap C) = \{1, 4\}$

$$\{1, 4\} = \{1, 4\}$$

Leyes de identidad (unión e intersección de conjuntos)

En el capítulo 1 citamos la unión y la intersección como operaciones entre conjuntos; ahora en este apartado resolveremos varios problemas citándolas como **leyes de identidad** y aplicando sus propiedades.

Sean los conjuntos:

$U = \{0, 4, 5, 7, 8, 9\}$

$A = \{4, 5, 9\}$

$B = \{0, 5, 8, 9\}$

a) La unión de un conjunto cualquiera con el conjunto vacío es igual al conjunto original.

$$A \cup \varnothing = A$$

Ejemplo:

■ **1.** Sea el conjunto A citado, obtener:

$A \cup \varnothing$

$\{4, 5, 9\} \cup \varnothing = \{4, 5, 9\}$

b) La unión de un conjunto cualquiera con el conjunto universal es igual al conjunto universal.

$$B \cup U = U$$

Ejemplo:

■ **1.** Sean los conjuntos B y U citados, determina:

$B \cup U = U$

$\{0, 5, 8, 9\} \cup \{0, 4, 5, 7, 8, 9\} = \{0, 4, 5, 7, 8, 9\}$

c) La intersección de cualquier conjunto con el conjunto universal es igual al conjunto original.

$$A \cap U = A$$

Ejemplo:

■ **1.** Sean los conjuntos A y U citados, determina:

$A \cap U = A$

$\{4, 5, 9\} \cap \{0, 4, 5, 7, 8, 9\} = \{4, 5, 9\}$

d) La intersección de cualquier conjunto con el conjunto vacío es igual al conjunto vacío.

$$B \cap \varnothing = \varnothing$$

Ejemplo:

■ **1.** Sea el conjunto B citado, determina:

$B \cap \varnothing = \varnothing$

$\{0, 5, 8, 9\} \cap \varnothing = \varnothing$

Leyes de complemento

Sean los conjuntos:

$U = \{a, b, c, d, e\}$

$A = \{a, c, d\}$

a) La unión de un conjunto con su complemento es igual al conjunto universal.

$$A \cup A' = U$$

Ejemplo:

■ **1.** Sean los conjuntos U y A citados, determina:

$$A \cup A' = U$$

$A' = \{a, b, c, d, e\} - \{a, c, d\} = \{b, e\}$

$A \cup A' = \{a, c, d\} \cup \{b, e\} = \{a, b, c, d, e\}$

$U = \{a, b, c, d, e\}$

$$\{a, b, c, d, e\} = \{a, b, c, d, e\}$$

b) La intersección de un conjunto con su complemento es igual al conjunto vacío.

Ejemplo:

■ **1.** Sean los conjuntos U y A citados, determina:

$A \cap A' = \varnothing$

$A' = \{a, b, c, d, e\} - \{a, c, d\} = \{b, e\}$

$A \cap A' = \{a, c, d\} \cap \{b, e\} = \varnothing$

$\varnothing = \varnothing$

c) El doble complemento de un conjunto es igual al conjunto original.

$$(A')' = A$$

Ejemplo:

■ **1.** Con los conjuntos U y A citados, determina:

$(A')' = A$

$A = \{a, c, d\}$

$U = \{a, b, c, d, e\}$

$A' = \{b, e\}$

$(A')' = \{a, c, d\} = A$

d) El complemento del conjunto universal es igual al conjunto vacío.

$$U' = \varnothing$$

Ejemplo:

■ **1.** Con el conjunto U citado, comprueba:

$$U' = \varnothing$$

$$\{a, b, c, d, e\} - \{a, b, c, d, e\} = \varnothing$$

$$\varnothing = \varnothing$$

e) El complemento del conjunto vacío es igual al conjunto universal.

$$\varnothing' = U$$

Ejemplo:

1. Con el conjunto U citado, verifica que $\varnothing' = U$:

$$\varnothing' = U$$

$\varnothing' = \{a, b, c, d, e\} - \varnothing = \{a, b, c, d, e\}$ | $\{a, b, c, d, e\} = \{a, b, c, d, e\} = U$

Leyes de De Morgan

Las leyes de De Morgan relacionan la unión y la intersección de conjuntos.

Primera ley

El complemento de la unión de dos conjuntos es igual a la intersección de los complementos de cada uno.

$$(A \cup B)' = A' \cap B'$$

Ejemplo:

1. Con los conjuntos:

$U = \{0, 4, 5, 7, 8, 9\}$

$A = \{4, 5, 9\}$

$B = \{0, 5, 8, 9\}$

Primero calculamos $(A \cup B)$ y luego el complemento de $A \cap B$.

$A \cup B = \{0, 4, 5, 8, 9\}$

$(A \cup B)' = \{0, 4, 5, 7, 8, 9\} - \{0, 4, 5, 8, 9\} = \{7\}$

A continuación obtenemos $A' \cap B'$:

$A' = \{0, 4, 5, 7, 8, 9\} - \{4, 5, 9\} = \{0, 7, 8\}$

$B' = \{0, 4, 5, 7, 8, 9\} - \{0, 5, 8, 9\} = \{4, 7\}$

$A' \cap B' = \{0, 7, 8\} \cap \{4, 7\} = \{7\}$

Entonces tenemos que:

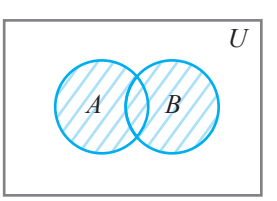
$(A \cup B)' = A' \cap B'$

$\{7\} = \{7\}$

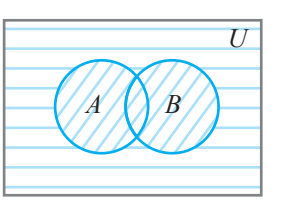
Representación gráfica de la primera ley de De Morgan con diagramas de Venn:

$(A \cup B)' = A' \cap B'$

Primero representamos $(A \cup B)'$

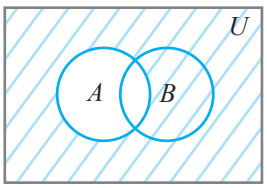


$(A \cup B)$

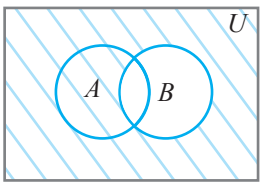


$(A \cup B)'$

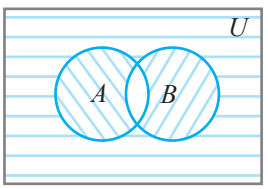
Observa las rayas horizontales que representan el conjunto $(A \cup B)'$.



A'



B'



Observa las rayas horizontales que representan $A' \cap B'$.

Segunda ley

El complemento de la intersección de dos conjuntos es igual a la unión de los complementos de cada uno.

$$(A \cap B)' = A' \cup B'$$

Ejemplo:

1. Con los conjuntos citados al principio de este párrafo calcula $(A \cap B)'$.

Primero calculamos:

$(A \cap B) = \{4, 5, 9\} \cap \{0, 5, 8, 9\} = \{5, 9\}$

$(A \cap B)' = \{0, 4, 5, 7, 8, 9\} - \{5, 9\} = \{0, 4, 7, 8\}$

A continuación obtenemos:

$A' \cup B'$

$A' = \{0, 4, 5, 7, 8, 9\} - \{4, 5, 9\} = \{0, 7, 8\}$

$B' = \{0, 4, 5, 7, 8, 9\} - \{0, 5, 8, 9\} = \{4, 7\}$

$A' \cup B' = \{0, 7, 8\} \cup \{4, 7\} = \{0, 4, 7, 8\}$

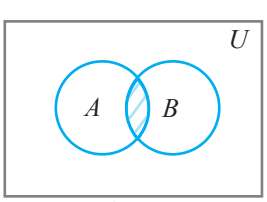
Entonces tenemos que:

$(A \cap B)' = A' \cup B'$

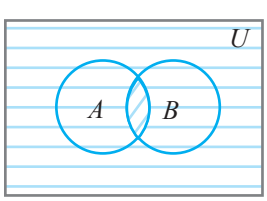
$\{0, 4, 7, 8\} = \{0, 4, 7, 8\}$

Representación gráfica de la segunda ley de De Morgan con diagramas de Venn.

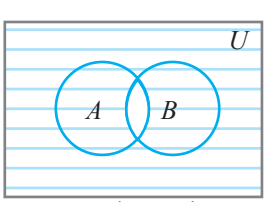
Primero expresamos $(A \cap B)'$



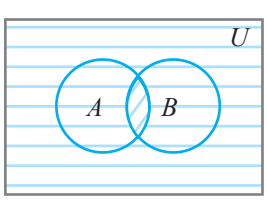
$A \cap B$



$(A \cap B)'$



A' y B'



$A' \cup B'$

Observa las rayas horizontales.

Leyes de la teoría de conjuntos

Leyes de idempotencia	$A \cup A = A$	$A \cap A = A$
Leyes asociativas	$(A \cap B) \cap C = A \cap (B \cap C)$	$(A \cup B) \cup C = A \cup (B \cup C)$
Leyes conmutativas	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Leyes distributivas	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Leyes de identidad	$A \cup \varnothing = A$ $A \cup U = U$	$A \cap U = A$ $A \cap \varnothing = \varnothing$
Leyes de complemento	$A \cup A' = U$ $A \cap A' = \varnothing$ $(A')' = A$	
Leyes de De Morgan	$(A \cup B)' = A' \cap B'$	$(A \cap B)' = A' \cup B'$

Teorema: $A - B = A \cap B'$

Algunos autores identifican esta última expresión como la **ley de la diferencia de dos conjuntos** y la definen así:

La diferencia de dos conjuntos es igual a la intersección del minuendo con el complemento del sustraendo.

Ejemplo:

1. Con los siguientes conjuntos:

$U = \{0, 4, 5, 7, 8, 9\}$

$A = \{4, 5, 9\}$

$B = \{0, 5, 8, 9\}$

verifica que $A - B = A \cap B'$

$A - B = \{4, 5, 9\} - \{0, 5, 8, 9\} = \{4\}$

$B' = \{0, 4, 5, 7, 8, 9\} - \{0, 5, 8, 9\}$

$B' = \{4, 7\}$

$A \cap B' = \{4, 5, 9\} \cap \{4, 7\} = \{4\}$

$\{4\} = \{4\}$

Problemas resueltos

1. En una fiesta infantil hay tres sabores de agua fresca: guayaba, naranja y tamarindo. Representa con diagramas de Venn y con expresiones matemáticas los siguientes sucesos:

a) Ningún niño consume agua de guayaba.

b) A ninguno le gustan los tres sabores disponibles.

c) Prefieren sólo agua de guayaba.

d) Prefieren agua de guayaba o de naranja, pero no de tamarindo.

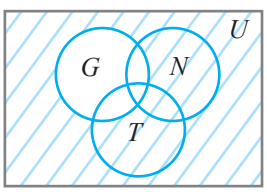
Solución:

Sean G = Guayaba

N = Naranja

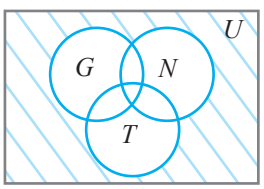
T = Tamarindo

a)



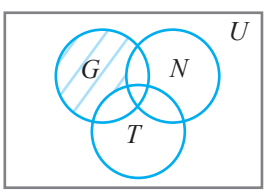
G'

b)



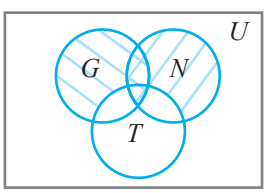
$(G \cup N \cup T)'$

c)



$G - [(G \cap N) \cup (G \cap T)]$

d)



$(G \cup N) - T$

■ **2.** Una orquesta de 20 músicos decide formar dos grupos musicales, uno de clásica y otro de música de salón. El primer grupo lo integran 8 personas y el segundo 12 personas. Si tres de los músicos pertenecen a los dos grupos, ¿cuántos miembros de la orquesta original decidieron no pertenecer a ningún grupo?

Solución:

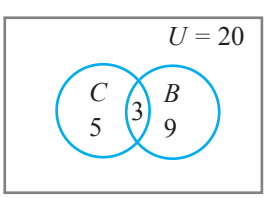
$U = 20$

$n(C) = 8$ músicos en música clásica

$n(B) = 12$ músicos en música de salón

$n(X)$ = músicos que no pertenecen a ningún grupo

La expresión *pertenecen a los dos grupos* sugiere la intersección de los conjuntos.



$$n(U) = n(C) + n(B) - n(C \cap B) + n(X)$$

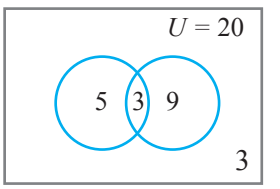
Sustituimos:

$20 = 8 + 12 - 3 + n(x)$

$20 - 17 = n(x)$

$n(x) = 3$

Tres músicos decidieron no pertenecer a ningún grupo.



3. El departamento de personal de la maquiladora HANDEX necesita contratar programadores, 25 de ellos realizarán tareas de programación de sistemas y 40 ocuparán el área de desarrollo de programas de aplicación; 10 de todos los contratados deben realizar trabajos de ambas especialidades ¿cuántos programadores se deben contratar en total?

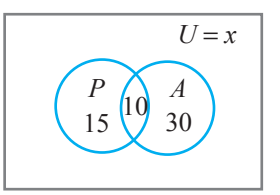
Solución:

$n\,(U) = x$

$n\,(P) = 25$ programadores de sistemas

$n\,(A) = 40$ desarrollo de programas de aplicación

La expresión "ambas especialidades" sugiere la intersección de los conjuntos.



$$n\,(U) = n\,(P) + n\,(A) - n\,(P \cap A)$$

Sustituimos:

$n\,(U) = 25 + 40 - 10$

$n\,(U) = 55$

La empresa deberá contratar 55 programadores, que es la cardinalidad del conjunto U.

4. ¿A cuántas amas de casa se entrevistaron en una encuesta para conocer sus preferencias sobre los programas de televisión si se obtuvieron los siguientes datos?

19 películas; 23 conciertos; 17 noticieros

Algunas personas agregaron otras preferencias a los resultados:

9 películas y conciertos, únicamente.

6 conciertos y noticieros, únicamente.

4 películas y noticieros, únicamente.

3 películas, conciertos y noticieros.

Solución:

$n\,(U) = x$

$n\,(P) = 19$ películas

$n\,(C) = 23$ conciertos

$n\,(N) = 17$ noticieros

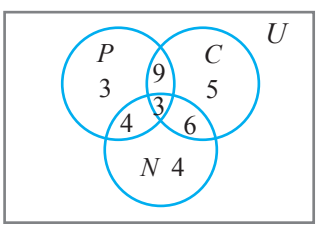
$n\,(P \cap C) - n\,(P \cap C \cap N) = 9$ películas y conciertos,

$n\,(C \cap N) - n\,(P \cap C \cap N) = 6$ conciertos y noticieros,

$n\,(P \cap N) - n\,(P \cap C \cap N) = 4$ películas y noticieros,

$n\,(P \cap N \cap C) = 3$ películas, noticieros y conciertos

La encuesta también arrojó la siguiente información: de las 19 que prefieren películas, a 9 de ellas además les gustan los conciertos, lo que sugiere la intersección de los conjuntos películas y conciertos:



Escribimos el número 3 en la intersección de los conjuntos películas, conciertos y noticieros; escribimos el número 9 en la intersección de películas y conciertos; escribimos el número 4 en la intersección de películas y noticieros y, por último, escribimos el número 6 en la intersección de conciertos y noticieros.

De las 19 personas que prefieren películas sumamos $9 + 3 + 4 = 16$, número que restamos a 19 y queda $19 - 16 = 3$, cifra que anotamos dentro del conjunto P. Del mismo modo procedemos con los conjuntos de conciertos C y noticieros N.

Para obtener la cardinalidad del conjunto universal $n\,(U)$ sumamos la cardinalidad de cada conjunto y le restamos los valores de las intersecciones, desarrollo que se expresa así:

$n\,(U) = n\,(P) + n\,(\mathrm{C}) + n\,(N) - n\,(P \cap N) - n\,(N \cap C) - n\,(P \cap C) + n\,(P \cap C \cap N)$

Sustituimos:

$n\,(U) = 19 + 23 + 17 - 7 - 9 - 12 + 3 = 59 - 25$

$n\,(U) = 34$

Se entrevistaron 34 amas de casa.

Nota: Una vez que aprendas y aceptes el razonamiento que permite obtener un resultado, no será necesario expresar la solución con la notación de conjuntos, pues observando directamente la representación gráfica con diagramas de Venn y con sumas y restas aritméticas se obtiene el resultado.

5. En una encuesta que se realizó entre 1 500 trabajadores y padres de familia de una empresa se obtuvieron los datos siguientes: 775 tienen casa propia, 800 automóvil, 760 servicio de televisión por cable. De todas estas personas, 300 señalaron que además de tener casa tenían automóvil, 250 casa y cable, 270 automóvil y cable y 200 personas, de mejor situación económica, tenían las tres cosas. ¿Cuántos tienen sólo dos cosas? ¿Cuántos al menos dos? ¿Cuántos padres de familia no tienen ninguno de estos tres bienes?

Solución:

$n\,(U) = 1\,500$

$n\,(C) = 775$ tienen casa

$n\,(A) = 800$ automóvil

$n\,(V) = 760$ cable

$n\,(C \cap A) - n\,(A \cap C \cap V) = 300$ casa y automóvil

$n\,(C \cap V) - n\,(A \cap C \cap V) = 250$ casa y cable

$n\,(A \cap V) - n\,(A \cap C \cap V) = 270$ automóvil y cable

$n\,(A \cap C \cap V) = 200$ las tres cosas

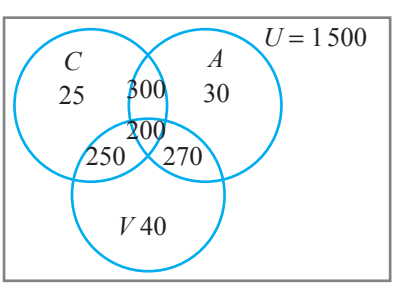
Se quiere saber cuántos:

a) Sólo tienen dos cosas.

b) Tienen al menos dos cosas.

c) Tienen menos de dos cosas.

d) Carecen de los tres bienes.



a) Sólo tienen dos cosas:

$n(C \cap A) + n(C \cap V) + n(A \cap V)\ 3\ [n\,(A \cap C \cap V)]$

Sustituimos:

$500 + 450 + 470 - 3(200) = 820$

b) Tienen al menos dos cosas:

$n(C \cap A) + n(C \cap V) + n(A \cap V) - 2\ [n\,(C \cap A \cap V)]$

Sustituimos:

$500 + 450 + 470 - 2(200) = 1\,020$

c) Tienen menos de dos cosas (aquí se incluyen las personas que no tienen nada):

$n[C - (C \cap A) - (C \cap V) + (C \cap A \cap V)]$

$+ n[A - (A \cap C) - (A \cap V) + (A \cap C \cap V)]$

$+ n[V - (V \cap A) - (V \cap C) + (A \cap C \cap V)]$

$+ n\,(A \cup B \cup C)'$

Sustituimos:

$775 - 500 - 450 + 200$

$+ 800 - 500 - 470 + 200$

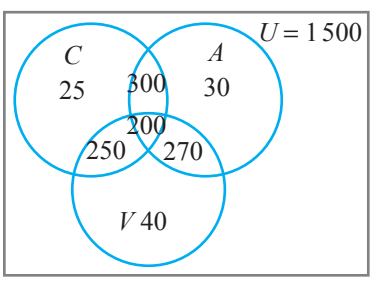
$+ 760 - 470 - 450 + 200$

$+ 385 = 480$

d) No tienen casa, ni automóvil ni cable:

Observamos la gráfica del diagrama de Venn y obtenemos:

$1\,500 - 25 - 300 - 200 - 250 - 30 - 270 - 40 = 1\,500 - 1\,115 = 385$



Solución:

820 sólo tienen dos cosas.

1 020 al menos dos cosas.

480 menos de dos cosas.

385 ninguno de los bienes citados.

6. En una escuela de enseñanza media superior, los alumnos que reprobaron matemáticas, física o química tendrán que presentar examen extraordinario, mientras que los alumnos que reprobaron las tres materias deberán repetir el curso. Analiza los siguientes resultados:

8% aprobaron las tres materias

28% aprobaron matemáticas y física

24% aprobaron matemáticas y química

36% aprobaron física y química

56% aprobaron matemáticas

59% aprobaron física

56% aprobaron química

¿Qué porcentaje de alumnos deberá repetir el curso?

¿Qué porcentaje aprobó sólo una materia?

Solución:

M = {alumnos que aprobaron matemáticas} = 56%

F = {alumnos que aprobaron física} = 59%

Q ={(alumnos que aprobaron química} = 56%

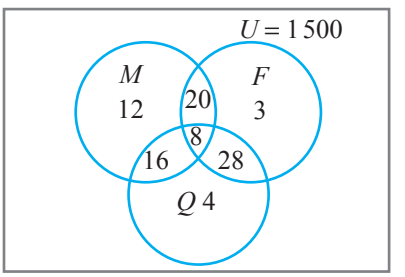
$M \cap F \cap Q = 8\%$

$M \cap F = 28\%$

$M \cap Q = 24\%$

$F \cap Q = 36\%$

Representamos los datos en un diagrama de Venn:



Solución:

Los alumnos que aprobaron alguna materia es la suma de todas las cantidades indicadas:

$12 + 20 + 3 + 16 + 8 + 28 + 4 = 91\%$

Porcentaje de alumnos que deben repetir el curso: $100 - 91 = 9\%$

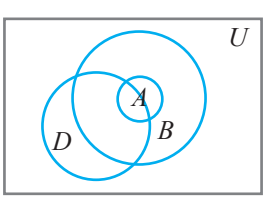
Porcentaje de alumnos que sólo aprobó una materia: $12 + 3 + 4 = 19\%$

¡Aplícate!

Traza con diagramas de Venn, en un conjunto universal U y tres conjuntos no vacíos A, B y D, de manera que tengan las características que se indican en cada uno.

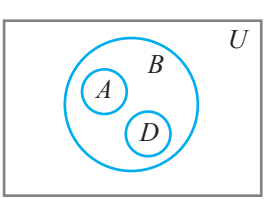
1. $A \subset B, A \cap D \neq \varnothing, D \not\subset B$

Sol.



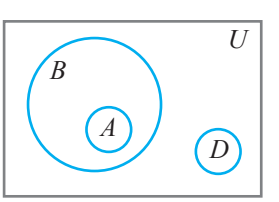
2. $A \subset B, A \cap D = \varnothing, D \subset B$

Sol.



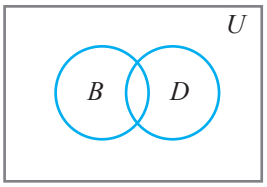
3. $A \subset B, B \cap D = \varnothing$

Sol.

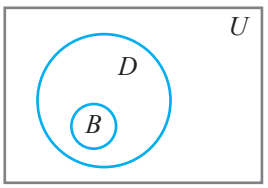


Con los diagramas de Venn siguientes, donde U es el conjunto universal, señala:

4. $B \cap D$

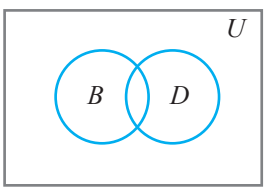


$B \cap D$

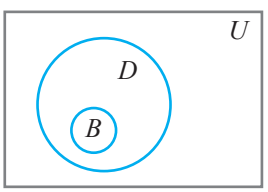


$B \cap D$

5. $D - B$

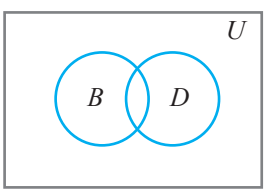


$D - B$

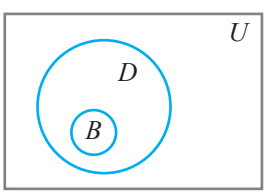


$D - B$

6. D'

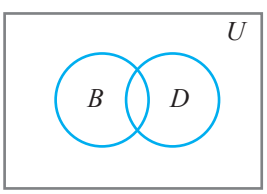


D'

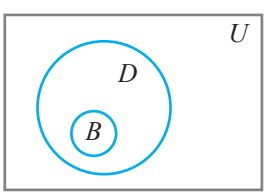


D'

7. $B' \cup D$

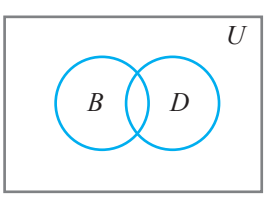


$B' \cup D$



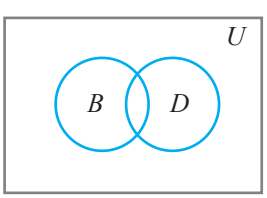
$B' \cup D$

8. $B \cap D'$



$B \cap D'$

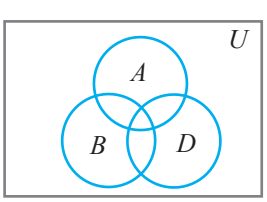
9. $B' - D'$



$B' - D'$

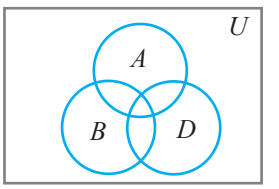
Con los conjuntos universal U y los conjuntos A, B, D, señala:

10. $A \cup (B - D)$



$A \cup (B - D)$

11. $(A \cup B) - (A \cup D)$

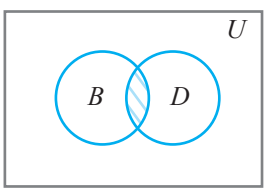


$(A \cup B) - (A \cup D)$

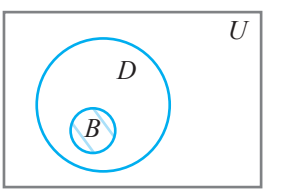
Soluciones a los ejercicios 4–11

■ Ejercicio 4

Sol.



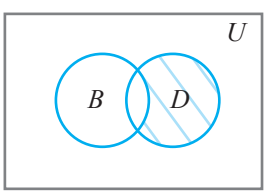
$B \cap D$



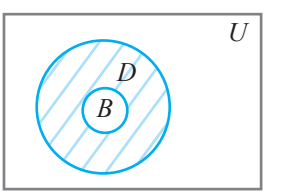
$B \cap D$

■ Ejercicio 5

Sol.



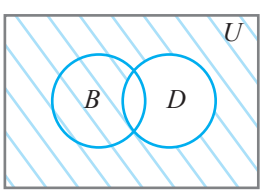
$D - B$



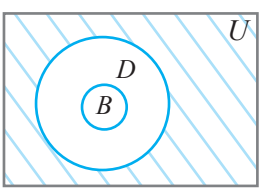
$D - B$

■ Ejercicio 6

Sol.



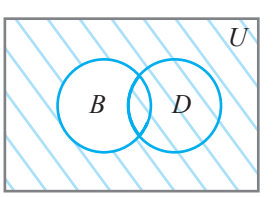
D'



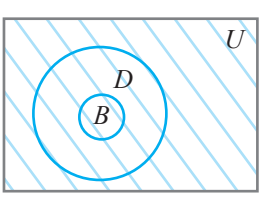
D'

■ Ejercicio 7

Sol.



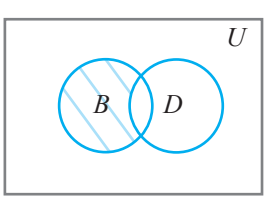
$B' \cup D$



$B' \cup D$

■ Ejercicio 8

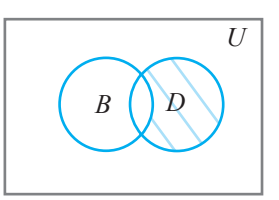
Sol.



$B \cap D'$

■ Ejercicio 9

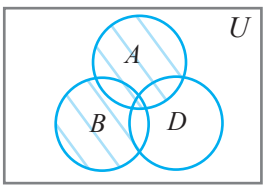
Sol.



$B' - D'$

■ Ejercicio 10

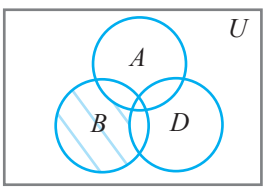
Sol.



$A \cup (B - D)$

■ Ejercicio 11

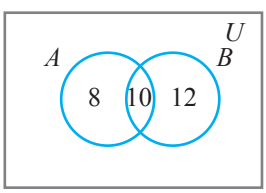
Sol.



$(A \cup B) - (A \cup D)$

Ejercicios de repaso

1. De acuerdo con el diagrama, calcula la cardinalidad de los conjuntos que se piden.



$n(U) = 50$

$n(A) =$

$n(B) =$

$n(A \cup B) =$

$n(A \cap B) =$

$n(A \cup B)' =$

Sol. $n(A) = 8 + 10 = 18$

$n(B) = 10 + 12 = 22$

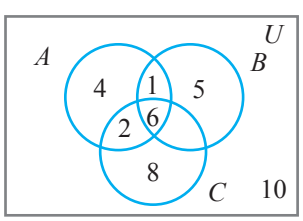
$n(A \cup B) = 8 + 10 + 12 = 30$

$n(A \cap B) = 10$

$n(A \cup B)' = n(U) - n(A \cup B)$

$= 50 - 30 = 20$

2. De acuerdo con el diagrama, calcula la cardinalidad de los conjuntos que se piden.



$n(U) =$

$n(A \cap B) =$

$n(B \cap C) =$

$n(A \cap B \cap C) =$

$n(A \cap C) =$

Sol. $n(U) = 4 + 1 + 5 + 2 + 6 + 8 + 10 = 36$

$n(A \cap B) = 1 + 6 = 7$

$n(B \cap C) = 6$

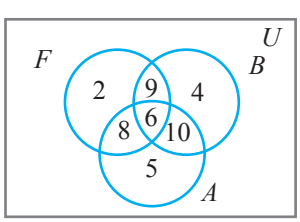
$n(A \cap B \cap C) = 6$

$n(A \cap C) = 2 + 6 = 8$

3. Bajo qué condiciones se cumple $n(A \cup B) = n(A) + n(B)$.

Sol. $A \cap B = \varnothing$

De acuerdo con el siguiente diagrama contesta las preguntas 4–8.



F = Personas a las que les gusta jugar fútbol.

B = Personas a las que les gusta jugar básquetbol.

A = Personas a las que les gusta jugar ajedrez.

4. ¿A cuántas personas les gusta jugar fútbol? ***Sol.*** $n(F) = 2 + 9 + 6 + 8 = 25$

5. ¿A cuántas personas les gusta jugar básquetbol? ***Sol.*** $n(B) = 9 + 4 + 10 + 6 = 29$

6. ¿A cuántas personas les gusta jugar ajedrez? ***Sol.*** $n(A) = 8 + 6 + 10 + 5 = 29$

7. ¿A cuántas personas les gusta jugar exactamente dos actividades?

Sol. $n(F \cap B) + n(B \cap A) + n(F \cap A) =$
$-3(F \cap B \cap A) = 15 + 14 + 16 - 3(6) = 27$

8. ¿A qué cantidad de personas les gusta jugar más de una actividad?

Sol. A los que les gusta jugar 2 actividades
+ a los que les gusta jugar tres actividades = $27 + 6 = 33$

9. Construye un diagrama de Venn–Euler en el que se represente el siguiente caso.

En un centro de lenguas:

a) No hay quien hable tres idiomas y una lengua indígena.

b) Tres personas hablan español, francés e inglés.

c) Diez personas hablan español y francés.

d) Dieciocho personas hablan español y alguna lengua.

e) Tres personas hablan francés y alguna lengua.

f) Dos personas hablan francés, alguna lengua y español.

g) Dieciséis personas hablan francés e inglés.

h) Dieciocho personas hablan en español e inglés.

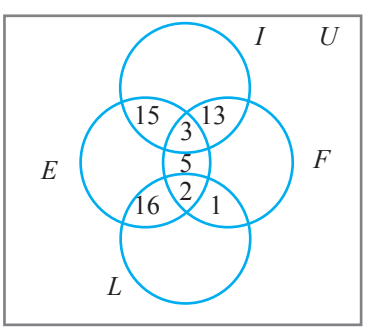
E = Español,

I = Inglés,

F = Francés,

L = Lenguas indígenas.

Sol.



10. En el ejercicio anterior, calcula la población total si no hay personas que hablan sólo un idioma o sólo una lengua.

Sol. $15 + 3 + 13 + 5 + 16 + 2 + 1 = 55$

11. En el ejercicio 9, ¿cuál es el idioma que más personas hablan en el centro de lenguas?

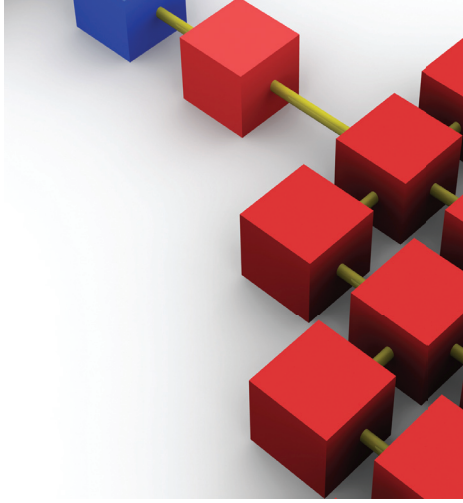
Sol. $n(I) = 15 + 3 + 13 = 31$

$n(E) = 15 + 3 + 5 + 2 + 16 = 41$

$n(F) = 13 + 3 + 5 + 2 + 1 = 24$

$n(L) = 16 + 2 + 1 = 19$

Es español



Capítulo 3

Análisis combinatorio

Introducción

Con frecuencia se presentan problemas en los que, por ejemplo, una institución bancaria tiene que dar a sus clientes una tarjeta de crédito o débito; una compañía de teléfonos debe asignar a cada suscriptor un número o bien, un gobierno estatal debe emitir una placa de circulación para vehículos particulares.

La solución de este tipo de problemas implica calcular cuántos subconjuntos distintos se pueden formar con un conjunto de números. Sin embargo, es importante que el sistema que se seleccione tenga suficiente amplitud para cubrir el número de usuarios previstos.

A cada número, objeto o suceso se le llama **elemento**; a cada colección o grupo de elementos se le identifica como una **combinación** y a cada ordenación única dentro de una combinación se le llama **permutación**.

Una *combinación* es un conjunto de elementos diferentes en cualquier orden.

La *permutación* se caracteriza por el orden de los elementos que la forman.

Todos los *elementos* de un conjunto pueden ser empleados en cualquier combinación o permutación, pero no es necesario utilizarlos a todos; generalmente estos elementos son de la misma especie aunque esta condición no es absolutamente necesaria.

Conceptos clave

- Elemento
- Combinación
- Permutación
- Principio multiplicativo
- Principio aditivo
- Factorial
- Permutaciones lineales
- Permutaciones circulares
- Combinaciones
- Coeficiente binomial
- Triángulo de Tartaglia

Principios fundamentales del conteo

Al estudiar las operaciones con conjuntos nos referimos al *conjunto producto* señalando que dados dos conjuntos A y B:

$$A \times B = \{(a, b) \mid a \in A, b \in B\}$$

A este conjunto de *parejas ordenadas* se le llama *producto cartesiano* de A y B, y es el principio fundamental para contar, el cual citamos a continuación:

Si un conjunto finito A contiene n elementos y un conjunto B también finito contiene r elementos, entonces hay nr parejas ordenadas donde $a \in A$ y $b \in B$; el resultado del producto $A \times B$ contiene nr elementos.

Este principio puede extenderse a cualquier número de conjuntos y aplicarse a muchas situaciones de conteo.

Principio multiplicativo

Si un primer suceso, el que algunos autores citan como *evento*, puede formarse de p_1 maneras diferentes, entonces, dos sucesos pueden verificarse siguiendo el orden indicado de $p_1\, p_2$ maneras diferentes.

Problemas

1. ¿De cuántas maneras diferentes se pueden seleccionar parejas de distinto sexo de un grupo de 4 hombres y 6 mujeres?

Solución:

Como cada hombre puede ser seleccionado de cuatro maneras diferentes y cada mujer puede ser seleccionada de 6 maneras diferentes, cada pareja puede ser escogida de:

$$4(6) = 24 \text{ maneras diferentes}$$

Si el suceso incluye más de dos sucesos diferentes podemos ampliar el principio multiplicativo, de manera que, después de haber ocurrido los dos primeros sucesos, puede ocurrir un tercero de p_3 maneras diferentes, un cuarto de p_4 maneras diferentes, y por último un n–ésimo de p_n maneras diferentes. Así, los sucesos pueden ocurrir en el orden siguiente:

$$p_1\, p_2\, p_3\, p_4 \ldots, p_n \text{ maneras diferentes.}$$

2. Cuántos números naturales nones existen que tengan una expresión numérica (numeral) de tres dígitos con los elementos: {0, 1, 2, 3, 4, 5, 6, 7}

Solución:

Para razonar este tipo de problemas, resulta útil elaborar diagramas como los que se presentan a continuación. Analiza el procedimiento para hacerlos:

Traza pequeñas cuadrículas | | | | | o coloca una serie de rayas pequeñas _ _ _ _, tantas como sean necesarias.

La cifra de las centenas es cualquiera de los siete elementos {1, 2, 3, 4, 5, 6, 7}. Consideramos únicamente siete porque no podemos escoger el cero para las centenas. Después escribimos un 7 en el primer espacio.

7		

La cifra de las decenas es cualquiera de los ocho elementos {0, 1, 2, 3, 4, 5, 6, 7}, así que escribimos el 8 en el segundo espacio.

7	8	

La cifra de las unidades debe ser cualquier número non de los elementos citados, que son {1, 3, 5, 7}. Luego escribimos el 4 en el tercer espacio.

7	8	4

Por el principio multiplicativo, la solución es:

$7(8)(4) = 224$ son los números naturales nones que tienen tres cifras y que se pueden expresar con los elementos {0, 1, 2, 3, 4, 5, 6, 7}.

3. Con los dígitos del 0 al 9 se quieren formar números de 4 cifras sin repetir cifras en ninguno de los números formados.

a) ¿Cuántos se pueden formar?

b) ¿Cuántos números son impares?

c) ¿Cuántos números son divisibles entre 2?

d) ¿Cuántos números son mayores o iguales que 3 000?

Solución:

Tenemos {0, 1, 2, 3…, 9}. Son diez elementos.

a) | 9 | 9 | 8 | 7 |

9(9) (8) (7) = 4 536

Se pueden formar 4 536 números.

b) | 8 | 8 | 7 | 5 |

Para la última cifra y para formar números impares en que pusimos el 5, tenemos {1, 3, 5, 7, 9}, que son 5 elementos:

8(8) (7) (5) = 2 240

Se pueden formar 2 240 números impares.

c) Total de números obtenidos – números impares = números pares.

4 536 – 2 240 = 2 296

d) | 7 | 9 | 8 | 7 |

Para un número mayor o igual que 3 000 están el 3 001, 3 002; por lo tanto, la primera cifra se puede escoger de {3, 4, 5, 6, 7, 8, 9}, que son 7 elementos.

7(9) (8) (7) = 3 528

Se pueden obtener 3 528 números mayores o iguales que 3 000.

4. Calcula cuántos números enteros de tres cifras se pueden obtener con los dígitos 2, 3, 5, 7 en los casos siguientes:

a) No se permite la repetición de las cifras en ninguno de los números.

b) Se permite la repetición de las cifras en los números.

Solución:

a) No se permite la repetición.

Tenemos tres lugares para llenar: | | | |

El primero se puede llenar de cuatro formas diferentes usando cualquiera de los números 2, 3, 5, 7; en este caso, $p_1 = 4$; para el segundo caso y dado que sólo quedan 3 números disponibles, tenemos $p_2 = 3$ y $p_3 = 2$; por lo tanto:

$$4(3)\ (2) = 24$$

Los números enteros de 3 cifras que pueden formarse con los dígitos 2, 3, 5, 7 y sin repetir cifras son 24.

b) Se permite la repetición.

Tenemos tres lugares para llenar: □□□

El primer, segundo y tercer lugar se pueden llenar, cada uno, con 4 números; por lo tanto:

$$4(4)\,(4) = 64$$

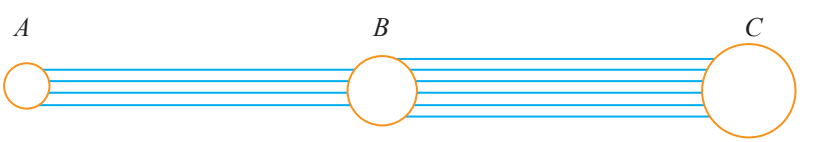
Los números enteros de 3 cifras que pueden formarse con los dígitos 2, 3, 5, 7 y con repetición son 64.

5. Para llegar de la ciudad A a la ciudad B hay 4 caminos. A su vez, para llegar de la ciudad B a la ciudad C hay 6 caminos. Si todos los caminos son diferentes, de cuántas formas es posible:

a) Viajar de la ciudad A hasta la C pasando por la B.

b) Hacer el viaje redondo desde la ciudad A hasta la C, pasando por la B.

c) Hacer el viaje redondo desde la ciudad A hasta la C, pasando por la B pero sin utilizar el mismo camino más de una vez.



Solución:

a) $A \to B$ $B \to C$

4 formas 6 formas

$4(6) = 24$ formas diferentes

b) $A \to B$ $B \to C$ $C \to B$ $B \to A$

4 formas 6 formas 6 formas 4 formas

$4(6)\,(6)\,(4) = 576$ formas diferentes

c) $A \to B$ $B \to C$ $C \to B$ $B \to A$

4 formas 6 formas 5 formas 3 formas

$4(6)\,(5)\,(3) = 360$ formas diferentes

Principio aditivo

Cuando un suceso (evento) puede realizarse de dos maneras diferentes excluyentes (una o la otra) y la primera de ellas puede realizarse de p_1 maneras diferentes y la siguientes de p_2 maneras diferentes, entonces el suceso puede realizarse de $p_1 + p_2$ maneras diferentes.

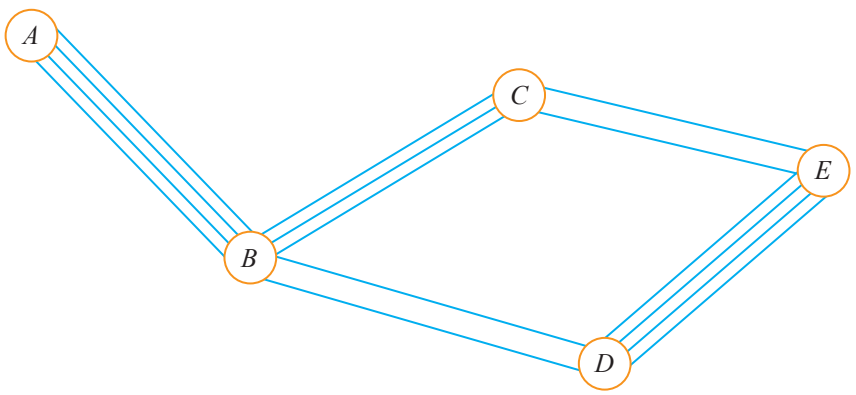
Este principio, al igual que el principio multiplicativo, puede generalizarse para procesos que incluyen tres o más operaciones excluyentes.

Analiza la situación siguiente:

En un problema se multiplican los sucesos cuando al primero le sigue (sucede) otro y se suman los sucesos cuando éstos son aislados.

Problemas

1. Cinco ciudades se comunican según el diagrama que presentamos a continuación:



De cuántas formas es posible:

a) Viajar desde A hasta E.

b) Hacer el viaje redondo desde A hasta E.

c) Hacer el viaje redondo desde A hasta E sin usar el mismo camino más de una vez.

Solución:

a) Viajar desde A hasta E.

Pasando por C

$A \rightarrow B$	$B \rightarrow C$	$C \rightarrow E$
4 formas	3 formas	2 formas

Por el principio multiplicativo:

$4(3)(2) = 24$ formas

Pasando por D

$A \rightarrow B$	$B \rightarrow D$	$D \rightarrow E$
4 formas	2 formas	4 formas

Por el principio multiplicativo:

$4(2)(4) = 32$ formas

El viaje desde A hasta E, ya sea pasando por C o por D, por el principio aditivo, se puede realizar en:

$24 + 32 =$ 56 formas

b) Viaje redondo desde A hasta E.

Por el principio multiplicativo obtenemos:

$A \to B \to C \to E$ y $E \to C \to B \to A$

[4(3)(2)][2(3)(4)] = 24(24) = 576 formas

$A \to B \to C \to E$ y $E \to D \to B \to A$

Sustituimos:

[4(3)(2)] [4(2)(4)] = 24(32) = 768 formas

$A \to B \to D \to E$ y $E \to D \to B \to A$

Sustituimos:

[4(2)(4)] [4(2)(4) = 32(32) = 1 024 formas

$A \to B \to D \to E$ y $E \to C \to B \to A$

Sustituimos:

[4(2)(4)] [2(3)(4)] = 32(24) = 768 formas

El viaje redondo desde A hasta E, por el principio aditivo se puede realizar en:

576 + 768 + 1 024 + 768 = 3 136 formas

c) Viaje redondo desde A hasta E sin usar el mismo camino más de una vez.

Por el principio multiplicativo obtenemos:

$A \to B \to C \to E$ y $E \to C \to B \to A$

Sustituimos:

[4(3)(2)] [1(2)(3)] = 24(6) = 144 formas

$A \to B \to C \to E$ y $E \to D \to B \to A$

Sustituimos:

[4(3)(2)] [4(2)(3)] = 24(24) = 576 formas

$A \to B \to D \to E$ y $E \to C \to B \to A$

Sustituimos:

[4(2)(4)] [2(3)(3)] = 32(18) = 576 formas

$A \to B \to D \to E$ y $E \to D \to B \to A$

Sustituimos:

[4(2)(4)] [3(1)(3)] = 32(9) = 288 formas

El viaje redondo desde *A* hasta *E*, sin usar el mismo camino más de una vez y por el principio aditivo, se puede realizar en:

144 + 576 + 576 + 288 = 1 584 formas diferentes

2. Calcula cuántos números impares menores que l0 000 pueden expresarse usando los números 0, 3, 6, 9.

Solución:

Como en el resultado se tendrán números de una, dos, tres o cuatro cifras podemos considerar cada caso por separado; y para que sean impares, de los números 0, 3, 6, 9 únicamente podemos disponer del 3 y del 9. Recuerda que el cero no puede ser cifra inicial de ninguna expresión numérica de un número natural.

Números impares de una cifra: el 3 y el 9.

Números impares de dos cifras: 2(3) = 6 números.

Números impares de tres cifras: 2(4) (3) = 24 números.

Números impares de cuatro cifras: 2(4) (4) (3) = 96 números.

Los números impares menores que 10 000 que pueden expresarse usando los números 0, 3, 6, 9 por el principio aditivo son:

2 + 6 + 24 + 96 = 128 números

Conclusión: Muchos problemas de conteo se pueden resolver aplicando únicamente los principios multiplicativo y aditivo, pero en algunos casos esto implica un considerable número de operaciones y razonamientos de alta complejidad. Para evitarlos, se han obtenido fórmulas para situaciones específicas que facilitan la solución.

Factorial

El producto de cualquier número entero positivo *n* por todos los números enteros menores que *n* se llama *factorial de n* y se expresa con el símbolo *n*! Así, tenemos:

0! = 1 Por definición

1! = 1 (1) = 1

2! = 2 (1) = 2

3! = 3 (2) (1) = 6

4! = 4 (3) (2) (1) = 24

5! = 5 (4) (3) (2) (1) = 120

.

.

.

$n! = (n)\ (n-1)\ (n-2),\ldots,\ (1)$

El factorial de los primeros números enteros positivos se puede obtener con una calculadora común. Para números mayores, se puede utilizar la fórmula aproximada de Stirling o bien, consultando tablas elaboradas con resultados.

Permutaciones

Cuando el problema de un conteo consiste en ordenar los elementos de un conjunto, decimos que se trata de una permutación del conjunto.

Las permutaciones se clasifican en:

- **Lineales.** Pertenecen a este tipo los casos en los cuales los elementos se ordenan en una fila.
- **Circulares.** Son todos aquellos casos en que los elementos deben ser colocados alrededor de un círculo en una secuencia cíclica o cerrada.

Permutaciones lineales

a) Permutaciones de n elementos diferentes en grupos de r elementos.

b) Permutaciones de n elementos no todos diferentes entre sí.

c) Permutaciones de n elementos diferentes en grupos de n elementos.

En algunos casos, el número de lugares es menor que el número de elementos; sin embargo, también se pueden presentar casos en que se disponga de un mayor número de lugares que de elementos. Cuando esto ocurre, se permutan los lugares con los elementos.

Una permutación se expresa así:

$$nPr$$

pero también se pueden emplear las notaciones $P(n, r)$ y P^{n}_{r}, donde n es el total de elementos por acomodar y r es la cantidad de elementos que se toman de n.

El número de permutaciones de n elementos diferentes tomados en grupos de r en r se obtiene con la fórmula:

$$nPr = n(n-1)\,(n-2)...(n-r+1), \text{ donde } r \leq n$$

Demostración de fórmula

El resultado de nPr es igual al número total de formas en que pueden llenarse r lugares con n elementos diferentes: el primer lugar puede llenarse de n formas diferentes porque en este punto todos los elementos están disponibles. El segundo lugar puede llenarse de $n-1$ formas diferentes con los $n-1$ elementos restantes.

Asimismo, el tercer lugar se puede llenar de $n-2$ formas diferentes, y así sucesivamente.

Continuando con este proceso, observamos que el lugar r puede llenarse con:

$$n-(r-1) = n-r+1 \text{ formas diferentes.}$$

De donde, el número total de formas está dado por la fórmula citada.

$$nPr = n(n-1)\,(n-2)...(n-r+1), \text{ donde } r \leq n.$$

Si se multiplica el segundo miembro de esta igualdad por $\dfrac{(n-r)!}{(n-r)!}$

$$nPr = \frac{n(n-1)(n-2)...(n-r+1)(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

Obtenemos la fórmula: $nPr = \frac{n!}{(n-r)!}$ si los n elementos son diferentes.

Para obtener el número total de permutaciones de n objetos tomamos de n en n elementos. En la fórmula anterior hacemos $r = n$, de donde:

$$nPr = \frac{n!}{(n-n)!} = \frac{n!}{0!}$$ como 0! = 1 por definición, tenemos:

$$nPr = n!$$

Problemas

1. ¿Cuántas diferentes quintas de baloncesto pueden formarse con siete jugadores disponibles para jugar cualquier posición?

Solución:

El problema se puede resolver aplicando el principio fundamental del conteo o con la fórmula

$nPr = n(n-1)\,(n-2)\ldots\,(n-r+1)$

Sustituimos:

${}_7P_5 = 7(6)(5)(4)(3) = 2\,520$

Se pueden formar 2 520 quintas diferentes con 7 jugadores disponibles.

Observa cómo obtuvimos el último factor:

Sustituimos en $n - r + 1$

con $n = 7$

$r = 5$

$n - r + 1 = 7 - 5 + 1 = 3$, que es el último término que anotamos.

2. En una empresa, cinco ejecutivos asisten a una junta donde hay siete sillas. Calcula de cuántas formas pueden ocupar las sillas.

Solución:

Como únicamente se ocupan cinco sillas, el número de diferentes modos de ocuparlas es igual al número de permutaciones de siete objetos considerados en grupos de cinco. Esto se expresa así: ${}_7P_5$.

Con la fórmula $nPr = \frac{n!}{(n-r)!}$

Sustituimos con:

$n = 7$

$r = 5$

$$P_5 = \frac{7!}{(7-5)!} = \frac{7(6)(5)(4)(3)(2!)}{2!} = 2\,520$$

Las sillas se pueden ocupar de 2 520 formas.

■ **3.** Determina cuántos números de cinco cifras se pueden formar con los dígitos l, 2, 3, 4 y 5 sin repetir ningún dígito.

Solución:

Con la fórmula $nPr = n!$

Sustituimos con:

$n = 5$

$_5P_5 = 5! = 5(4)(3)(2)(1) = 120$

Se pueden formar 120 números.

■ **4.** Determina los siguientes valores:

a) $_7P_3$

b) $_{13}P_1$

c) $_4P_4$

Solución:

En cada ejercicio aplicamos la fórmula $nPr = \dfrac{n!}{(n-r)!}$

a) $_7P_3 = \dfrac{7!}{(7-3)!} = \dfrac{7!}{4!} = \dfrac{7(6)(5)(\cancel{4!})}{\cancel{4!}} = 210$

b) $_{13}P_1 = \dfrac{13!}{(13-1)!} = \dfrac{13!}{12!} = \dfrac{13(\cancel{12!})}{\cancel{12!}} = 13$

c) $_4P_4 = \dfrac{4!}{(4-4)!} = \dfrac{4!}{0!} = \dfrac{4!}{1} = 4(3)(2)(1) = 24$

Permutaciones de *n* elementos, no todos diferentes entre sí

Si n representa el número de permutaciones distintas de n elementos tomados de n en n, donde hay un tipo de n_1 elementos iguales entre sí; n_2 elementos iguales entre sí de un segundo tipo; n_3 elementos iguales entre sí de un tercer tipo y así sucesivamente hasta el grupo g que tiene ng elementos iguales entre sí, entonces el número de permutaciones de los n elementos considerados en grupos de n se obtiene con la siguiente fórmula:

$$P = \frac{n!}{n_1!\, n_2!\, n_3! \ldots n_g!}$$

Si en los n–elementos, n_1 son iguales, entonces n_2, n_3, n_g también son iguales.

Problemas

■ **1.** Calcula el número de permutaciones diferentes que pueden formarse con las letras de la palabra *matemáticas* tomadas a la vez.

Solución:

La solución de este problema implica obtener el número de permutaciones de 11 letras consideradas en un grupo de 11, donde la letra *a* aparece repetida 3 veces; la letra *t*, 2 veces y la letra *m*, 2 veces.

Con la fórmula $P = \dfrac{n!}{n_1!\, n_2!\, n_3! \ldots n_g!}$

Sustituimos con:

$n = 11$

$n_1 = 3$

$n_2 = 2$

$n_3 = 2$

$$P = \frac{11!}{3!\ 2!\ 2!} = \frac{11(\overset{5}{\cancel{10}})(9)(\overset{4}{\cancel{8}})(7)(6)(5)(4)(\cancel{3})(\cancel{2})(\cancel{1})}{\cancel{3}(\cancel{2})(\cancel{1})\ \cancel{2}\ (1)\quad \cancel{2}\ (1)}$$

$$= 11\ (5)\ (9)\ (4)\ (7)\ (6)\ (5)\ (4)$$

$$= 1\,663\,200$$

Se pueden formar 1 663 200 palabras, no necesariamente de uso común.

2. Calcula el número de permutaciones diferentes que pueden formarse con las letras de la palabra *acacia*, tomadas todas a la vez.

Solución:

Con la fórmula $P = \dfrac{n!}{n_1!\, n_2!\, n_3! \ldots n_g!}$

Sustituimos con:

$n = 6$

$n_1 = 3$

$n_2 = 2$

$$P = \frac{6!}{3!\ 2!} = \frac{(6)(5)(\overset{2}{\cancel{4}})(\cancel{3})(\cancel{2})(\cancel{1})}{\cancel{3}(\cancel{2})(\cancel{1})\quad \cancel{2}\ (\cancel{1})}$$

$$= 6\ (5)\ (2)$$

$$= 60$$

Se pueden formar 60 palabras, no necesariamente de uso común.

Permutaciones circulares (cíclicas)

Pertenecen a este tipo todos aquellos casos en que n elementos diferentes deben ser colocados alrededor de un círculo en una secuencia cíclica o cerrada. También en este caso interviene la totalidad de los elementos disponibles.

El número de permutaciones circulares de n objetos donde la colocación del primer elemento puede considerarse como fija y los $n - 1$ elementos restantes pueden arreglarse de $(n - 1)!$ formas, se expresa así:

$$PC = (n - 1)!$$

Problemas

1. Tres mujeres y tres hombres se van a sentar de modo que sus lugares queden alternados. Calcula de cuántas formas pueden hacerlo si:

a) Se sientan en línea recta.

b) Se sientan en una mesa circular.

Solución:

a) Se sientan en línea recta.

Podemos considerar que las mujeres se sientan en los lugares con número impar y los hombres en los lugares con número par.

Esto puede hacerse de 3! y 3! formas diferentes. El mismo resultado se obtiene si las mujeres se sientan en lugares par y ellos en lugares impar. Así tenemos:

$2\,(3!\,3!) = 2\,(6)\,(6) = 72$

Si se sientan en línea recta lo pueden hacer en 72 formas.

Solución:

b) Se sientan en una mesa circular.

Primero podemos sentar a las mujeres alrededor de la mesa en 2! (el primer elemento está fijo), entonces quedan 3 lugares alternados para sentar a los hombres, que pueden sentarse en 3! formas, de donde:

$2!\,3! = 2\,(6) = 12$

Si se sientan en una mesa circular lo pueden hacer en 12 formas.

2. Calcula de cuántas formas se pueden sentar 8 personas alrededor de una mesa si:

a) Pueden sentarse en cualquier forma.

b) Si dos personas determinadas no deben estar una al lado de la otra.

Solución:

Con la fórmula $PC = (n - 1)!$

a) Sustituimos con:

$n = 8$

$PC = (8 - 1)! = 7! = 7\,(6)\,(5)\,(4)\,(3)\,(2)\,(1)$

$= 5\,040$

Las 8 personas pueden sentarse en 5 040 formas.

En este problema hemos considerado que una de las personas se puede sentar en cualquier lugar y las siete restantes pueden sentarse en 7!

b) Consideremos a las dos personas que no pueden estar juntos como si fueran una sola; entonces tenemos siete personas que pueden sentarse en círculo de 6! formas, y las dos personas consideradas como una

sola pueden ordenarse entre sí de 2! formas; por lo tanto, el número de ordenaciones es de:

$PC = 6!\ 2! = 6\ (5)\ (4)\ (3)\ (2)\ (1)\ (2)(1)$

$= 1\,440$

Como en **a)** tenemos el número total de formas en que siete personas se pueden sentar alrededor de una mesa y ahora dos de ellas no deben quedar juntas, entonces:

$5\,040 - 1\,440 = 3\,600$ es la solución

Combinaciones

Cuando el problema de un conteo consiste en obtener, en cualquier orden, grupos de r elementos de un total disponible de n elementos diferentes, es una combinación.

El número de combinaciones de n elementos tomados de r en r se expresa nCr, algunos autores emplean también los símbolos $C(n, r)$, C_r^n y $\binom{n}{r}$. Esta última expresión se lee n sobre r.

En las expresiones anteriores:

n es el total de elementos disponibles.

r el tamaño de los equipos a elegir.

La expresión $\binom{n}{r}$ se conoce como **coeficiente binomial**.

El número de combinaciones de n elementos diferentes tomados de r en r está dado por la fórmula:

$$\binom{n}{r} = nCr = \frac{n(n-1)\ldots(n-r+1)}{r!}, \text{ donde } r \leq n$$

Demostración de la fórmula

De cada combinación de r elementos diferentes podemos formar $r!$ permutaciones; por lo tanto, de todas las combinaciones se pueden formar un total de $nCr \cdot r!$

Si esta expresión la igualamos con nPr, que es la representación de una permutación de n elementos diferentes tomados de r en r, tenemos:

$$nCr \cdot r! = nPr$$

Despejando:

$$nCr = \frac{nPr}{r!}$$

y como $nPr = \dfrac{n!}{(n-r)!}$ si sustituimos en la expresión anterior el valor de nPr, queda la fórmula:

$$nCr = \frac{n!}{(n-r)!r!}$$

Si usamos el coeficiente binomial obtenemos:

$$\binom{n}{r} = nCr = \frac{n!}{(n-r)!r!}$$

A. Combinaciones que se forman con un conjunto.

Problema

■ **1.** Calcula de ${}_8C_5$, ${}_7C_3$ y de ${}_9C_4$.

Solución:

Como la fórmula $\binom{n}{r} = nCr = \frac{n!}{(n-r)!r!}$

Para ${}_8C_5$:

$$\binom{8}{5} = {}_8C_5 = \frac{8!}{(8-5)!5!} = \frac{8!}{3!5!} = \frac{8(7)(6)(5!)}{3(2)(1)(5!)} = 56$$

Para ${}_7C_3$

$$\binom{7}{3} = {}_7C_3 = \frac{7!}{(7-3)!3!} = \frac{7!}{4!3!} = \frac{7(6)(5)(4!)}{4!3!} = 35$$

Para ${}_9C_4$:

$$\binom{9}{4} = {}_9C_4 = \frac{9!}{(9-4)!4!} = \frac{9!}{5!4!} = \frac{9(8)(7)(6)(5!)}{5!4(3)(2)(1)} = 126$$

B. Combinaciones que se forman con varios conjuntos.

Problemas

■ **1.** En una maquiladora se presentan a solicitar trabajo 9 hombres y 5 mujeres. ¿De cuántas formas el jefe de personal puede hacer la selección si únicamente puede contratar a 6 hombres y a 3 mujeres?

Solución:

Con la fórmula

$$\binom{n}{r} = nCr = \frac{n!}{(n-r)!r!}$$

Hacemos dos sustituciones con:

$n = 9$

$r = 6$

y con:

$n = 5$

$r = 3$

Por el principio multiplicativo del conteo:

$$_9C_6 \cdot {}_5C_3 = \frac{9!}{(9-6)!6!} \cdot \frac{5!}{(5-3)!3!}$$

$$= \frac{\overset{3}{9}(8)(7)(\cancel{6!})}{\cancel{3}(\cancel{2})(\cancel{1})(\cancel{6!})} \cdot \frac{\overset{2}{5}(4)(\cancel{3!})}{\cancel{2}(\cancel{1})(\cancel{3!})} = 840$$

Se pueden contratar a los 6 hombres y a las 3 mujeres de 840 formas.

■ **2.** Un alumno de enseñanza media superior tiene 7 libros de física y 5 de matemáticas. Calcula de cuántas maneras se pueden ordenar 3 libros de física y 2 de matemáticas en un librero.

Solución:

Posibles combinaciones de libros de física:

$$_7C_3 = \frac{7!}{(7-3)!3!} = \frac{7(6)(5)(\cancel{4!})}{\cancel{4!}(3)(2)(1)} = \frac{210}{6} = 35$$

Posibles combinaciones de libros de matemáticas:

$$_5C_2 = \frac{5!}{(5-2)!2!} = \frac{5(4)(\cancel{3!})}{\cancel{3!}2(1)} = \frac{20}{2} = 10$$

Posibles maneras de elegir los 5 libros (principio multiplicativo del conteo):

$_7C_3 \cdot {}_5C_2 = 35\ (10) = 350$

Cada una de estas 350 combinaciones posibles para colocar 5 libros diferentes es de $_5P_5 = 5!$, cada combinación se puede ordenar de 5! maneras diferentes:

$5! = (5)\ (4)\ (3)\ (2)\ (1) = 120$

Hay 120 permutaciones para cada combinación; por lo tanto, el número total de permutaciones es de:

$350\ (120) = 42\,000$

Hay 42 000 formas de ordenar los libros, 3 de física y 2 de matemáticas en un librero de un total de 7 libros de física y 5 de matemáticas.

Relaciones de las permutaciones y las combinaciones

Algunas relaciones de las permutaciones y las combinaciones son las siguientes:

A. De las permutaciones

■ **1.** $nPn - 1 = nPn$

Demostración:

Con las fórmulas $nPr = \frac{n!}{(n-r)!}$ y $nPn = n!$

si $r = (n - 1)$

$$nPn-1=\frac{n!}{[n-(n-1)]}=\frac{n!}{n-n+1}=\frac{n!}{1}=n!$$

Como:

$nPn=n!$

Por la propiedad transitiva de la igualdad:

$nPn-1=nPn$

■ **2.** $nPn-r=\left(\frac{1}{r}\right)!\,(nPn)$

Demostración:

Con las fórmulas $nPr=\frac{n!}{(n-r)!}$ y $nPn=n(n-1)\,(n-2)...1=n!$

si $nPn-r=\frac{n!}{[n-(n-r)]!}=\frac{n!}{(n-n+r)!}=\frac{n!}{r!}$

Como:

$$\left(\frac{1}{r}\right)!\,(nPn)=\left(\frac{1}{r}\right)!\,(n!)=\frac{n!}{r!}$$

Por la propiedad transitiva de la igualdad:

$$nPn-r=\left(\frac{1}{r}\right)!\,(nPn)$$

■ **3.** $nPn-2=\frac{1}{2}(nPn)$

Demostración:

Con las fórmulas $nPr=\frac{n!}{(n-r)!}$ y $nPn=n(n-1)\,(n-2)...1=n!$

si $r=(n-2)$

$$nPn-2=\frac{n!}{[n-(n-2)]!}=\frac{n!}{2}$$

$$\frac{1}{2}(nPn)=\frac{n!}{2}$$

Por la propiedad transitiva de la igualdad:

$$nPn-2=\frac{1}{2}(nPn)$$

B. De las combinaciones

■ **1.** $\binom{n}{o}=1$; además $\binom{n}{n}=1$

Demostración:

$$\binom{n}{o}=\frac{n!}{0!n!}=\frac{n!}{n!}=1$$

Para:

$$\binom{n}{n}=\frac{n!}{(n-n)!\,n!}=\frac{1}{0!}=1$$

■ **2.** $\binom{n}{r}=\binom{n}{n-r}$

Demostración:

$$\binom{n}{r}=\frac{n!}{(n-r)!r!}$$

$$\binom{n}{n-r}=\frac{n!}{[n-(n-r)]!(n-r)!}=\frac{n!}{(n-r)!r!}$$

Por transitividad

$$\binom{n}{r}=\binom{n}{n-r}$$

■ **3.** $\binom{n}{1}=n$

Demostración:

$$\binom{n}{1}=\frac{n!}{1!(n-1)!}=\frac{n(\cancel{n-1})!}{1(\cancel{n-1})!}=n$$

$$\binom{n}{1}=n$$

■ **4.** $\binom{n}{r}+\binom{n}{r+1}=\binom{n+1}{r+1}$

Demostración:

$$\binom{n}{r}+\binom{n}{r+1}=\frac{n!}{r!(n-r)!}+\frac{n!}{(r+1)![n-(r+1)]!}$$

Multiplicamos el numerador y el denominador por $r+1$ y $n-r$, respectivamente en cada sumando.

$$=\frac{n!(r+1)}{(r+1)r!(n-r)!}+\frac{n!(n-r)}{(r+1)!(n-r)[n-(r+1)]!}$$

$$=\frac{n!(r+1)}{(r+1)!(n-r)!}+\frac{n!(n-r)}{(r+1)!(n-r)!}$$

$$=\frac{n!(r+1)+n!(n-r)}{(r+1)!(n-r)!}$$

$$=\frac{n!(\cancel{r}+1+n-\cancel{r})}{(r+1)!(n-r)!}$$

$$=\frac{n\,!(n+1)}{(r+1)!(n-r)!}=\frac{(n+1)!}{(r+1)!(n-r)!}=\binom{n+1}{r+1}$$

$$\binom{n}{r}+\binom{n}{r+1}=\binom{n+1}{r+1}$$

Estas propiedades de las combinaciones son las que permiten obtener el triángulo de Tartaglia, el cual facilita el cálculo de los coeficientes de las potencias de un binomio. Al final del siguiente capítulo veremos cómo se utiliza en el teorema del binomio.

Resumen

A. Principios fundamentales del conteo

- Principio multiplicativo: $p_1 \cdot p_2 \cdot p_3 ... p_n$

  Verificación de n sucesos en el que cada uno puede verificarse de p_i maneras diferentes.

- Principio aditivo: $p_1 + p_2 + p_3 ... p_n$

  Un suceso tiene n maneras diferentes de realizarse, excluyentes entre sí y cada una de ellas sucede de p_i maneras distintas.

B. Permutaciones

Fórmulas

$nPr = n(n-1)(n-2)...(n-r+1)$, donde $r \leq n$.

Permutaciones de n elementos tomados de r en r.

$$nPr = \frac{n!}{(n-r)!}$$

$nPn = n!$ permutaciones de n elementos.

$$P = \frac{n!}{n_1!\, n_2!\, n_3! \ldots n_g!}$$

Si en los n elementos y repeticiones, n_1 son iguales, n_2 son iguales… n_g son también iguales.

- Relaciones de las permutaciones:

$$nPn - 1 = nPn$$

$$nPn - r = \left(\frac{1}{r}\right)!(nPn)$$

$$nPn - 2 = \frac{1}{2}(nPn)$$

$$PC = (n-1)!$$

C. Combinaciones

Fórmulas

$$\binom{n}{r} = nCr = \frac{n(n-1)\ldots(n-r+1)}{r!}$$

$$\binom{n}{r} = nCr = \frac{n!}{(n-r)!r!}$$

Relaciones de las combinaciones:

$\binom{n}{o} = 1$ además $\binom{n}{n} = 1$

$$\binom{n}{r}=\binom{n}{n-r}$$

$$\binom{n}{1}=n$$

$$\binom{n}{r}+\binom{n}{r+1}=\binom{n+1}{r+1}$$

D. Orden

En una permutación importa el orden y se relaciona a sucesiones ordenadas: parejas ordenadas, triadas ordenadas, etcétera.

En las combinaciones no importa el orden y se relacionan con la selección de un subconjunto de un conjunto dado.

Problemas resueltos

En cada uno de los problemas siguientes decide si se trata de una permutación o de una combinación. Determina el resultado.

■ **1.** Calcula el número de palabras de 5 letras, no necesariamente con significado, que se pueden formar con 12 letras diferentes.

Solución:

Permutación

Con la fórmula $nPr=\dfrac{n!}{(n-r)!}$

Sustituimos con $n=12$, $r=5$

$$nPr=\frac{n!}{(n-r)!}=\frac{12!}{(12-5)!}=\frac{12!}{7!}=\frac{12(11)(10)(9)(8)(\cancel{7}!)}{\cancel{7}!}=95\,040$$

95 040 palabras.

■ **2.** Calcula el número de palabras, no necesariamente con significado, que pueden formarse seleccionando 6 consonantes y 2 vocales entre 10 consonantes y 4 vocales, todas diferentes.

Observa que en el problema anterior, el número de palabras a formar era de exactamente 5 letras, por eso lo identificamos como permutación; ahora las consonantes y las vocales se van a seleccionar de diferentes grupos.

Solución:

Combinación y permutación.

Con la fórmula $\binom{n}{r}=nCr=\dfrac{n!}{(n-r)!r!}$

Seleccionamos las consonantes.

Sustituimos con $n=10$ y $r=6$

$$\binom{10}{6} = {}_{10}C_6 = \frac{10!}{(10-6)!6!} = \frac{10!}{4!6!} = \frac{10(9)(8)(7)(\cancel{6}!)}{4!\cancel{6}!} = \frac{10(\overset{3}{\cancel{9}})(\cancel{8})(7)}{\cancel{4}(\cancel{3})(\cancel{2})(1)} = 210$$

En forma semejante podemos seleccionar para las vocales.

Sustituimos con $n = 4$ y $r = 2$

$$\binom{4}{2} = {}_4C_2 = \frac{4!}{(4-2)!2!} = \frac{4!}{2!2!} = \frac{\cancel{4}(3)(2)(1)}{\cancel{2}(1)\cancel{2}(1)} = 6$$

Con base en el principio multiplicativo tenemos que por cada una de las 210 formas para seleccionar las consonantes hay 6 formas de selección de las vocales; por lo tanto:

210 (6) = 1 260 formas

Efectuadas cada una de estas selecciones, las 8 letras pueden permutarse (consonantes y vocales) en 8! formas diferentes; por ello, el número total de palabras, por el principio multiplicativo es de:

1 260(8!) = 1 260 (8)(7)(6)(5)(4)(3)(2)(1) = 50 803 200 formas diferentes.

50 803 200 palabras.

3. Calcula de cuántas maneras diferentes pueden colocarse 7 libros en un librero.

Solución:

Permutación para ordenar 7 elementos tomados todos a la vez.

En la fórmula $nPn = n!$

Sustituimos $n = 7$

${}_7P_7 = 7! = 7(6)\ (5)\ (4)\ (3)\ (2)\ (1) = 5\,040$

Se pueden colocar en 5 040 formas diferentes.

4. ¿De cuántas maneras diferentes pueden colocarse 7 libros tomando sólo 3 de ellos a la vez?

Solución:

Permutación de 7 elementos tomados de 3 en 3.

En la fórmula $nPr = n(n-1)(n-2)...(n-r+1)$

Sustituimos con $n = 7$ y $r = 3$

${}_7P_3 = 7(7-1)\ (7-2)...(7-3+1)$

$= 7(6)\ (5) = 210$

Se pueden colocar en 210 formas.

Otra solución:

Aplicando la combinación y la permutación primero calculamos de cuántas maneras podemos seleccionar 3 libros de 7.

En la fórmula $\binom{n}{r} = nCr = \dfrac{n!}{(n-r)!r!}$

Sustituimos con $n = 7$ y $r = 3$.

$$_7C_3 = \frac{7!}{(7-3)!3!} = \frac{7!}{4!3!} = \frac{7(6)(5)(4!)}{4!(3)(2)(1)} = 35$$

$$_7C_3 = 35$$

Pero cada uno de ellos se puede presentar de 3! formas, entonces por el principio multiplicativo tenemos:

35 (3!) = 35 (3) (2) = 210 maneras distintas de colocar los libros.

Obtuvimos el mismo resultado.

5. ¿De cuántas maneras diferentes se puede integrar el comité de un club deportivo con un presidente, un secretario y un tesorero. El número total de socios es de 15 personas.

Solución:

Permutación de 15 elementos tomados 3 a la vez.

En la fórmula $nPr = n(n-1)(n-2)...(n-r+1)$

Sustituimos con $n = 15$ y $r = 3$

$$_{15}P_3 = 15(14)(13)$$

$$= 2\,730$$

Se pueden formar 2 730 comités.

Otra solución:

Aplicando la combinación y la permutación: inicialmente calculamos cuántos grupos diferentes de 3 elementos podemos formar con 15 personas.

En la fórmula $\binom{n}{r} = nCr = \dfrac{n!}{(n-r)!r!}$

Sustituimos con $n = 15$ y $r = 3$

$$\binom{15}{3} = {}_{15}C_3 = \frac{15!}{(15-3)!3!} = \frac{15!}{12!3!} = \frac{^{5}15^{7}(14)(13)(12!)}{12!(3)(2)(1)} = 455$$

$$_{15}C_3 = 455$$

En cada uno de los 455 grupos, los comités serán diferentes dependiendo de quién sea presidente, secretario y tesorero. Para cada grupo hay 3! ordenaciones distintas, entonces por el principio multiplicativo tenemos:

455 (3!) = 455 (3) (2) (1) = 2 730 comités.

Obtuvimos el mismo resultado.

6. Se va a organizar un comité de investigación de 5 personas entre 7 representantes de un partido mayoritario y 6 del minoritario. Calcula el número de comités que se pueden formar si deben constar:

a) Exactamente de 3 representantes del partido mayoritario.

b) Por lo menos 3 representantes del partido minoritario.

Solución:

a) Exactamente de 3 representantes del partido mayoritario.

$$_7C_3 = \frac{7!}{(7-3)!3!} = \frac{7!}{4!3!} = \frac{7(6)(5)(4!)}{4!3!} = \frac{7(6)(5)}{3(2)(1)} = 35$$

Del partido minoritario salen los 2 representantes para completar el comité de 5.

$$_6C_2 = \frac{6!}{(6-2)!2!} = \frac{6!}{4!2!} = \frac{6(5)(4!)}{4!2!} = \frac{30}{2} = 15$$

Por el principio multiplicativo, el número total de comités de 5 miembros es de:

$35\ (15) = 525$

b) Por lo menos tres representantes del partido minoritario.

En este caso tendremos tres tipos de comités:

1) Tres representantes del partido minoritario y dos del mayoritario.

2) Cuatro personas del partido minoritario y uno del mayoritario.

3) Cinco del partido minoritario y ninguno del mayoritario.

Solución:

1) Representantes del partido minoritario y dos del mayoritario.

$$_6C_3 = \frac{6!}{(6-3)!3!} = \frac{6(5)(4)(3!)}{3!(3)(2)} = 20$$

2 del mayoritario

$$_7C_2 = \frac{7!}{(7-2)!2!} = \frac{7(6)(5!)}{5!(2)(1)} = 21$$

Por el principio multiplicativo hay $20(21) = 420$ comités.

2) Selección de 4 elementos de 6 que hay en el partido minoritario.

$$_6C_4 = \frac{6!}{(6-4)!4!} = \frac{6(5)(4!)}{2!4!} = \frac{30}{2} = 15$$

Selección de 1 elemento de 7 que hay en el partido mayoritario.

$$_7C_1 = \frac{7!}{(7-1)!(1)!} = \frac{7(\not{6}!)}{\not{6}!(1)} = 7$$

Por el principio multiplicativo, hay 15(7) = 105 comités.

3) Selección de 5 de 6 personas del partido minoritario.

$$_6C_5 = \frac{6!}{(6-5)!5!} = \frac{6\,\not{5}!}{1!\not{5}!} = \frac{6}{1} = 6$$

Por el principio aditivo, el total de comités que incluyen por lo menos 3 representantes del partido minoritario son:

420 + 105 + 6 = 531

■ **7.** El gobierno de un estado decidió iniciar un programa de reemplacamiento para los autos particulares. Ahora las nuevas placas tendrán dos letras del alfabeto y tres dígitos (del 0 al 9). ¿Cuántas placas se fabricarán con estas características?

Solución:

Permutación.

Dado que no hay ninguna limitación se pueden usar cualquiera de las 28 letras del alfabeto y cualquiera de los dígitos del 0 al 9; por ejemplo, la placa AA 111 es un número permitido.

Además, se acepta que el dígito que debe ir a continuación de la segunda letra puede ser el cero. Primero elegimos la letra que ocupa el primer lugar en la placa, como hay 28 letras para escoger, tenemos para el primer lugar $n_1 = 28$, así 28 _ _ _ _ de igual manera para el segundo lugar $n_2 = 28$.

Como cualquiera de los número del 0 al 9 se pueden usar en cualquiera de los otros 3 lugares, tenemos $n_3 = 10$, $n_4 = 10$, $n_5 = 10$.

Los cinco lugares de la placa quedan así:

28 28 10 10 10.

Cuando al primer hecho le sucede otro, se multiplica. Así, el número total de placas es de:

28 (28) (10) (10) (10) = 784 000

■ **8.** Un funcionario de un banco decide que los números de las tarjetas de débito se cambien de manera que no se repitan las letras o los números de cada una, las cuales incluirán primero dos letras del alfabeto y luego cuatro números de los dígitos del 0 al 9. Calcula el número de tarjetas que se van a fabricar.

Solución:

Para determinar el número de tarjetas en que ninguna letra o número aparezcan repetidos, tenemos $n_1 = 28$, $n_2 = 27$, ya que para el segundo lugar sólo quedan 27 letras disponibles del alfabeto. En igual forma para los números $n_3 = 10$, $n_4 = 9$, $n_5 = 8$, $n_6 = 7$. Así: 28 27 10 9 8 7 .

Y como en un problema, cuando al primer hecho le sucede otro, se multiplica. Así, el número total de tarjetas es de:

$$28\,(27)\,(10)\,(9)\,(8)\,(7) = 3\,810\,240$$

Nota: En los problemas que siguen calculamos por separado el último término y no directamente en la fórmula.

■ **9.** Una bolsa contiene 6 bolas rojas numeradas del 1 al 6 y 8 bolas azules numeradas del 1 al 8. ¿De cuántos modos se pueden seleccionar 6 bolas de manera que 2 sean rojas y 4 azules?

Solución:

a) Combinaciones de 6 elementos tomados de dos en dos.

Con la fórmula $\binom{n}{r} = nCr = \dfrac{n\,(n-1)...(n-r+1)}{r!}$

Sustituimos con $n = 6$ y $r = 2$

Calculamos el último término con:

$n - r + 1$

$6 - 2 + 1 = 5$

$$\binom{6}{2} = {}_6C_2 = \frac{6(5)}{2(1)} = 15$$

b) Combinación de 8 elementos tomados de cuatro en cuatro con la fórmula anterior.

Sustituimos con $n = 8$ y $r = 4$

Calculamos el último término con:

$n - r + 1$

$8 - 4 + 1 = 5$

$$\binom{8}{4} = {}_8C_4 = \frac{\cancel{8}^{2}(7)(\cancel{6})(5)}{\cancel{4}(\cancel{3})(\cancel{2})(1)} = 70$$

Por cada una de las combinaciones de bolas rojas tenemos 70 combinaciones de bolas azules, entonces:

$15(70) = 1\,050$

Hay 1 050 formas diferentes en que se pueden sacar 6 bolas de manera que 2 sean rojas y 4 azules.

■ **10.** La carta de un restaurante indica que hay 4 sopas, 7 estilos de carne, 8 ensaladas y 5 postres. ¿De cuántos modos se puede ordenar una comida consistente en una sopa, un plato de carne, tres ensaladas y un postre?

Solución:

Con la relación de las combinaciones $\binom{n}{1} = n$

y la fórmula $\binom{n}{r} = nCr = \frac{n!}{(n-r)!\,r!}$

${}_4C_1 = 4$

${}_7C_1 = 7$

${}_8C_3 = \frac{8!}{(8-3)!3!} = \frac{8!}{5!3!} = \frac{8(7)(6)(5!)}{5!3(2)} = 56$

Por cada sopa de las cuatro disponibles hay 7 estilos de carne.

Total de sopas y carnes: 4 (7) = 28

Por cada combinación de estas 28 hay 56 maneras de combinarlas con las ensaladas: 28(56) = 1 568

Por cada comida, y son 1 568 combinaciones que ahora tenemos, hay 5 postres, entonces con los 5 postres tenemos 1 568 (5) = 7 840 comidas diferentes.

■ **11.** ¿Cuántos grupos de 5 o más personas pueden formarse con 10 personas?

Solución:

Con la fórmula $\binom{n}{r} = nCr = \frac{n\,(n-1)...(n-r+1)}{r!}$

$\binom{10}{5} = {}_{10}C_5 =$

Calculamos el último término:

$n - r + 1$

$10 - 5 + 1 = 6$

$\binom{10}{5} = {}_{10}C_5 = \frac{\overset{2}{(10)}(9)\overset{2}{(8)}(7)(6)}{5(4)(3)(2)(1)} = 252$

Para:

$\binom{10}{6} = {}_{10}C_6$

Calculamos el último término:

$n - r + 1$

$10 - 6 + 1 = 5$

$\binom{10}{6} = {}_{10}C_6 = \frac{\overset{3}{10}(9)(8)(7)(6)(5)}{6(5)(4)(3)(2)(1)} = 210$

Para:

$\binom{10}{7} = {}_{10}C_7 =$

Calculamos el último término:

$n - r + 1$

$10 - 7 + 1 = 4$

$$\binom{10}{7} = {}_{10}C_7 = \frac{10\overset{3}{(9)}\overset{4}{(8)}(7)(6)(5)(4)}{7(6)(5)(4)(3)(2)(1)} = 120$$

Para:

$$\binom{10}{8} = {}_{10}C_8 =$$

Calculamos el último término:

$n - r + 1$

$10 - 8 + 1 = 3$

$$\binom{10}{8} = {}_{10}C_8 = \frac{10(9)(8)(7)(6)(5)(4)(3)}{8(7)(6)(5)(4)(3)(2)(1)} = 45$$

Para:

$$\binom{10}{9} = {}_{10}C_9 =$$

Calculamos el último término:

$n - r + 1$

$10 - 9 + 1 = 2$

$$\binom{10}{9} = {}_{10}C_9 = \frac{10(9)(8)(7)(6)(5)(4)(3)(2)}{9(8)(7)(6)(5)(4)(3)(2)(1)} = 10$$

Para:

$$\binom{10}{10} = {}_{10}C_{10} = 1$$

Como son hechos aislados, por el principio aditivo:

$252 + 210 + 120 + 45 + 10 + 1 = 638$

Se pueden formar 638 grupos de 5 o más personas.

12. En una escuela de enseñanza media superior los alumnos de matemáticas presentan un examen que incluye 16 problemas para resolver 8 de ellos. ¿Cuántos exámenes diferentes de 8 problemas se pueden escoger de esos 16?

Solución:

Combinación de 16 elementos tomados de ocho en ocho.

Con la fórmula $\binom{n}{r} = {}_nC_r = \frac{n\,(n-1)...(n-r+1)}{r!}$

Sustituimos con $n = 16$ y $r = 8$

$$\binom{16}{8} = {}_{16}C_8$$

$$= \frac{16(16-1)(16-2)...(16-8+1)}{8!}$$

$$= \frac{\cancel{16}(\cancel{15})(\overset{2}{\cancel{14}})(13)(\overset{2}{\cancel{12}})(11)(10)(9)}{\cancel{8}(\cancel{7})(\cancel{6})(\cancel{5})(\cancel{4})(\cancel{3})(\cancel{2})(1)}$$

$$= 12\,870$$

Se pueden escoger 12 870 exámenes diferentes.

¡Aplícate!

1. Calcular el número de ordenaciones o permutaciones diferentes que pueden formarse con las letras *A*, *B*, *C*, *D*, *E*, *F* que contengan 3 letras cada una.

Sol. Se pueden formar 120 palabras, no necesariamente de uso.

2. Se ordena en una fila 5 banderas rojas, 3 azules y 2 blancas. ¿De cuántas maneras posibles se pueden ordenar?

Sol. Las banderas se pueden ordenar en 2 520 formas.

3. ¿Cuántos comités de 5 personas se pueden formar con un conjunto de 18 personas?

Sol. Se pueden formar 8 568 comités de 5 personas.

4. Calcular el número de comités de 3 personas que pueden formarse con un conjunto de 14.

Sol. 364 comités.

5. ¿Cuántos comités de 3 personas se pueden formar de un grupo de 12 profesores?

Sol. Se pueden formar 220 comités.

6. ¿De cuántas formas diferentes se pueden separar 20 personas en grupos de 4 para jugar dominó?

Sol. Son 4 845 formas diferentes.

7. ¿De cuántas maneras distintas se pueden ordenar las letras de la palabra *Papantla* y de la palabra *Apizaco*?

Sol. 3 360 formas de ordenar la palabra Papantla.

Sol. 2 520 formas de ordenar la palabra Apizaco.

8. El dominó es un juego que consta de 28 fichas que normalmente se juega entre 4 personas donde cada una recibe 7 fichas. Calcula el número de juegos diferentes que pueden tener cada jugador.

Sol. 1 184 040 juegos diferentes.

Ejercicios de repaso

1. ¿En qué momento es conveniente utilizar el principio multiplicador?

2. ¿Cuándo se debe usar el principio aditivo?

3. ¿Cuál es la diferencia esencial entre los dos principios?

4. ¿En dónde es menor el número de formas de ordenar objetos, en permutaciones lineales o en permutaciones circulares? Justificar la respuesta.

5. Explica la diferencia entre el resultado obtenido de utilizar permutaciones (*nPr*) y combinaciones (*nCr*).

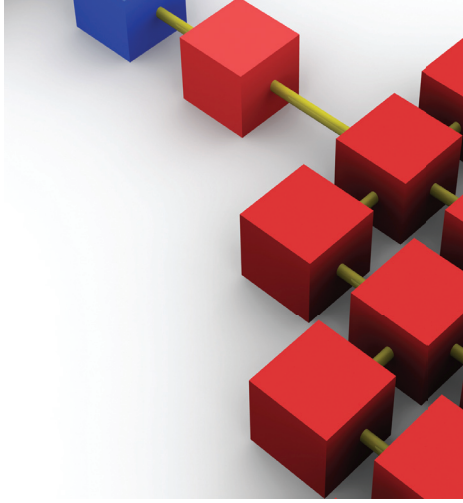
6. Una estudiante tiene 4 blusas, 5 faldas y 2 pares de zapatos. ¿Cuántas combinaciones diferentes se pueden realizar?

7. Hay dos caminos de *A* y *B*, tres de *C* a *B* y cinco de *A* a *C*. ¿Cuántas formas hay de salir de *B* a una ciudad inmediata?

8. ¿Cuántos números diferentes que contengan los dígitos 3, 8, 1 y 2 se pueden formar si deben ser de cuatro cifras?

9. ¿De cuántas formas es posible ordenar 6 libros en un lugar donde sólo hay espacio para tres?

10. ¿Cuántas maneras diferentes hay de ordenar las letras de la palabra *propositivo*?



Capítulo 4

Teorema del binomio
Triángulo de Tartaglia
Triángulo de Pascal

Teorema de binomio

En tu curso de aritmética y álgebra aprendiste que:

$a^0 = 1$

Por definición: Toda cantidad elevada a la cero potencia es igual a uno.

Ejemplos:

- **1.** $9^0 = 1$
- **2.** $10^0 = 1$
- **3.** $\left(\frac{1}{2}\right)^0 = 1$
- **4.** $\frac{d^m}{d^m} = d^0 = 1$ con $d \neq 0$
- **5.** $\frac{5^2}{5^2} = 5^0 = 1$

También aprendiste que:

- **Teorema del binomio.**

  Binomio a la n–ésima potencia.

Observa:

$(x+y)^1 = x+y$

$(x+y)^2 = x^2+2xy+y^2$

$(x+y)^3 = x^3+3x^2y+3xy^2+y^3$

$(x+y)^4 = x^4+4x^3y+6x^2y^2+4xy^3+y^4$

Conclusiones:

1. El número de términos es igual al grado del binomio más uno.
2. El grado del primer término es igual al grado del binomio y disminuye sucesivamente, en uno, en cada uno de los siguientes términos y es factor en todos los términos, menos en el último.
3. El segundo término aparece en el desarrollo con exponente uno y aumenta sucesivamente, en uno, en cada uno de los términos siguientes hasta llegar al exponente del binomio, el cual es el último del resultado.
4. El coeficiente del primer término del resultado es uno y el del segundo es el exponente del binomio; el último término también es uno.
5. El coeficiente de un término cualquiera es igual al coeficiente del término inmediato anterior por el exponente de x en ese término y dividido entre el número de términos desarrollados.
6. El grado de cada término es igual al grado del binomio.
7. Los términos que equidistan de los extremos tienen coeficientes iguales.
8. Cada término del binomio se considera con coeficiente y signo.

$$(x+y)^n = x^n + nx^{n-1}y + \frac{n(n-1)x^{n-2}y^2}{1(2)} + \frac{n(n-1)(n-2)x^{n-3}y^3}{1(2)(3)} + \ldots + y^n$$

Triángulo de Tartaglia

Las relaciones de las combinaciones ya citadas son:

a) $\binom{n}{0} = 1$, además $\binom{n}{n} = 1$

b) $\binom{n}{r} = \binom{n}{n-r}$

c) $\binom{n}{1} = n$

d) $\binom{n}{r} + \left(\frac{n}{r+1}\right) = \binom{n+1}{r+1}$

Permiten obtener el *triángulo de Tartaglia* que, junto con el triángulo de Pascal, facilitan el cálculo de los coeficientes de la potencia de un binomio.

$$\binom{0}{0}$$
$$\binom{1}{0}\quad\binom{1}{1}$$
$$\binom{2}{0}\quad\binom{2}{1}\quad\binom{2}{2}$$
$$\binom{3}{0}\quad\binom{3}{1}\quad\binom{3}{2}\quad\binom{3}{3}$$
$$\binom{4}{0}\quad\binom{4}{1}\quad\binom{4}{2}\quad\binom{4}{3}\quad\binom{4}{4}$$
$$\binom{5}{0}\quad\binom{5}{1}\quad\binom{5}{2}\quad\binom{5}{3}\quad\binom{5}{4}\quad\binom{5}{5}$$

$$1$$
$$1 \quad 1$$
$$1 \quad 2 \quad 1$$
$$1 \quad 3 \quad 3 \quad 1$$
$$1 \quad 4 \quad 6 \quad 4 \quad 1$$
$$1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1$$

Todos los números de los extremos son números uno.

Cada número que no sea de los extremos se obtiene sumando los dos superiores inmediatos con base en la propiedad

$$\binom{n}{r}+\left(\frac{n}{r+1}\right)=\binom{n+1}{r+1}$$

Triángulo de Pascal

Si necesitamos escribir la siguiente fila en el triángulo de Pascal, podemos hacerlo si ponemos cada término como la suma de los dos que están a su izquierda y a su derecha en la fila superior.

1
1 1
1 2 1
1 3 3 1
1 4 6 + 4 1
1 5 10 10 5 1

Problemas resueltos

Con la fórmula del binomio.

$$(x+y)^n = x^n + nx^{n-1}y + \frac{n(n-1)x^{n-2}y^2}{1(2)} + \frac{n(n-1)(n-2)x^{n-3}y^3}{1(2)(3)} + \ldots + y^n$$

Las relaciones de las combinaciones que permiten calcular el triángulo de Tartaglia.

La igualdad para obtener los coeficientes de $(x+y)^n$ es:

$$(x+y)^n = {}_nC_0\, x^n + {}_nC_1\, x^{n-1}y + {}_nC_2\, x^{n-2}y^2 + \ldots + {}_nC_n y^n$$

Por ejemplo, observa el triángulo de Tartaglia: en el cuarto renglón se citan los coeficientes para expresiones como $(2x+y)^4$ y que son $\binom{4}{0}\binom{4}{1}\binom{4}{2}\binom{4}{3}\binom{4}{4}$ y la expresión:

$${}_nC_{r-1}\, x^{n-(r-1)}y^{r-1}$$

para calcular el término que ocupa un lugar r que se obtiene sustituyendo los valores de n y de r.

Podemos resolver ejemplos como los siguientes:

Problemas

1. Calcula $(x+3y)^4$

$$(x+3y)^4 = \binom{4}{0}x^4 + \binom{4}{1}x^3(3y) + \binom{4}{2}x^2(3y)^2 + \binom{4}{3}x(3y)^3 + \binom{4}{4}(3y)^4$$

$$= 1x^4 + 4x^3(3y) + 6x^2(9y^2) + 4x(27y^3) + 1(81y^4)$$

$$= x^4 + 12x^3y + 54x^2y^2 + 108xy^3 + 81y^4$$

Los coeficientes de cada término los calculamos así:

$\binom{4}{0}=1$; $\binom{4}{1}=4$ relaciones de las combinaciones

$\binom{4}{2}=6$ con la fórmula $nCr=\dfrac{n!}{(n-r)!r!}$

$$\binom{4}{2}={}_4C_2=\frac{4!}{(4-2)!2!}=\frac{4!}{2!2!}=\frac{4(3)(2!)}{2!2!}=\frac{12}{2}=6$$

y toda la expresión (término) con:

$$6(x^2)(3y^2)=6x^2(9y)^2=54x^2y^2$$

$\binom{4}{3}=4$ lo obtuvimos con la fórmula $nCr=\dfrac{n!}{(n-r)!r!}$

$$\binom{4}{3}={}_4C_3=\frac{4!}{(4-3)!3!}=\frac{4!}{1!3!}=\frac{4(3)(2)(1)}{1(3)(2)(1)}=\frac{24}{6}=4$$

y toda la expresión (término) con:

$$4(x)(27y^3)=108xy^3$$

$\binom{4}{4}=1$ relación de las combinaciones

2. Determina el tercer término del desarrollo de $(x+3y)^4$

Sabemos por el resultado del problema anterior que es $54\,x^2y^2$

Solución:

El término que ocupa un lugar r se obtiene sustituyendo en la expresión:

$${}_nC_{r-1}\,x^{n-(r-1)}y^{r-1}$$

Con $n=4$ y $r=3$

$${}_4C_{3-1}\,x^{4-(3-1)}(3y)^{3-1}={}_4C_2x^2(3y)^2=6x^2(9y^2)$$

$$=54x^2y^2$$

3. Calcula $(x^2-3)^4$

Solución:

$$(x^2-3)^4$$

$$={}_4C_0(x^2)^4+{}_4C_1(x^2)^3(-3)+{}_4C_2(x^2)^2(-3)^2+{}_4C_3x^2(-3)^3+{}_4C_4(-3)^4$$

$$=1(x^8)+4x^6(-3)+6x^4(9)+4x^2(-27)+1(81)$$

$$=x^8-12x^6+54x^4-108x^2+81$$

4. Calcula $(2a + b^2)^6$

Solución:

Para resolver este producto y considerando que el triángulo de Tartaglia desarrollado en párrafos anteriores únicamente cita hasta la quinta potencia, a continuación damos los coeficientes de la sexta potencia.

$\binom{6}{0}\binom{6}{1}\binom{6}{2}\binom{6}{3}\binom{6}{4}\binom{6}{5}\binom{6}{6}$

$$(2a-b^2)^6=\binom{6}{0}(2a)^6+\binom{6}{1}(2a)^5(-b)+\binom{6}{2}(2a)^4(-b^2)^2+\binom{6}{3}(2a)^3(-b^2)^3$$

$$+\binom{6}{4}(2a)^2(-b^2)^4+\binom{6}{5}(2a)(-b^2)^5+\binom{6}{6}(-b^2)^6$$

$$=1(64a^6)+6(32a^5)(-b^2)+15(16a^4)(b^4)+20(8a^3)(-b^6)$$

$$+15(4a^2)(b^8)+6(2a)(-b^{10})+1(b^{12})$$

$$=64a^6-192a^5b^2+240a^4b^4-160a^3b^6+60a^2b^8-12ab^{10}+b^{12}$$

Los coeficientes de cada término los calculamos así:

$\binom{6}{0}=1$; $\binom{6}{1}=6$ relaciones de las combinaciones

$\binom{6}{2}=$ con la fórmula $nCr=\dfrac{n!}{(n-r)!r!}$, así:

$$\binom{6}{2}={}_6C_2=\frac{6!}{(6-2)!2!}=\frac{6(5)(4!)}{4!\,2(1)}=\frac{30}{2}=15$$

$$\binom{6}{3}={}_6C_3=\frac{6!}{(6-3)!3!}=\frac{6(5)(4)(3!)}{3!\,3!}=\frac{120}{6}=20$$

$$\binom{6}{4}={}_6C_4=\frac{6!}{(6-4)!4!}=\frac{6(5)(4!)}{2!\,4!}=\frac{30}{2}=15$$

$$\binom{6}{5}={}_6C_5=\frac{6!}{(6-5)!5!}=\frac{6(5!)}{1(5!)}=6$$

$\binom{6}{6}=1$ relaciones de las combinaciones

5. Determina el sexto término de $(2x + y)^9$

Solución:

El término que ocupa el lugar $r = 6$ se obtiene sustituyendo en la relación

$${}_nC_{r-1}\, x^{n-(r-1)} y^{r-1}$$

Sustituimos con $n = 9$ y $r = 6$

$${}_9C_{6-1}(2x)^{9-(6-1)} y^{6-1}$$

$${}_9C_5(2x)^4 y^5$$

A continuación, con la fórmula $nCr = \dfrac{n!}{(n-r)!\, r!}$ obtenemos el coeficiente de $16\, x^4y^5$.

$$\binom{9}{5} = \frac{9!}{(9-5)!5!} = \frac{9(8)(7)(6)(\cancel{5!})}{4!\cancel{5!}} = \frac{3\,024}{4(3)(2)(1)} = \frac{3\,024}{24} = 126$$

El sexto término es $126(16x^4y^5) = 2\,016x^4y^5$

6. Determina el quinto término de $(x + 2y)^7$

Solución:

El término que ocupa el lugar $r = 5$ se obtiene sustituyendo en la relación:

$${}_nC_{r-1}\, x^{n-(r-1)} y^{r-1}$$

Sustituimos con $n = 7$ y $r = 5$

$${}_7C_{5-1}\, x^{7-(5-1)}(2y)^{5-1}$$

$${}_7C_4\, x^3(2y)^4$$

A continuación, con la fórmula $nCr = \dfrac{n!}{(n-r)!\, r!}$ obtenemos el coeficiente de $x^3 16\, y^4$:

$$\binom{7}{4} = {}_7C_4 = \frac{7!}{(7-4)!4!} = \frac{7(6)(5)(\cancel{4!})}{3!\cancel{4!}} = \frac{210}{6} = 35$$

El quinto término es $35(x^3 16y^4) = 560\, x^3y^4$

Ejercicios de repaso

1. Dado el binomio $(x - 2y)^4$

a) ¿Cuántos términos tiene el resultado?

Sol. **a)** Tiene $n + 1$ términos. Como $n = 4$, el resultado tiene $4 + 1 = 5$ términos.

b) ¿Cómo se comportan los signos en el resultado?

Sol. **b)** Los signos se alteran: +, –, +, –, +. Esto en razón de las potencias del segundo término: $(-2y)^0$ es +, $(-2y)^1$ es –1, y así sucesivamente hasta $(-2y)^4$,que es +.

2. Dado el binomio $(-2x + 4)^5$

a) ¿Cuántos términos tiene el resultado?

Sol. **a)** Tiene $n + 1 = 5 + 1 = 6$ términos

b) ¿Cómo se comportan los signos en el resultado?

Sol. **b)** Los signos se alternan: –, +, –, +, –, +. Ello en razón del término primero: $(-2x)^5$ es –, $(-2x)^4$ es +, así sucesivamente hasta $(-2x)^0$, que es +.

c) Compruébalo desarrollando el binomio a la quinta.

Sol. **c)** $\binom{5}{0}(-2x)^5(4)^0+\binom{5}{1}(-2x)^4(4)^1+\binom{5}{2}(-2x)^3(4)^2+\binom{5}{3}(-2x)^2(4)^3+\binom{5}{4}(-2x)^1(4)^4$

$+\binom{5}{5}(-2x)^0(4)^5=-32x^5+5(16x^4)(4)+10(-8x)(16)+10(4x^2)(64)+(5)(-2x)(256)$

$1(1)(1\,024)-32x^5+320x^4-1\,280x^3+2\,560x^2-2\,560x+1\,024$

3. Resuelve $\left(\frac{1}{3}x+2y\right)^2$

Sol. $\left(\frac{1}{3}x+2y\right)^2=1\left(\frac{1}{3}x\right)^2+(2)\left(\frac{1}{3}x\right)(2y)+(2y)^2=\frac{1}{9}x^2+\frac{4}{3}xy+4y^2$

El primer factor de cada término se obtuvo del triángulo de Pascal. (1, 2, 1).

4. Calcula el coeficiente de x^0 en $(4x + 3y)^7$

Sol. x^0 pertenece al último término del resultado; entonces, se calcula:

$$\binom{7}{7}=1$$

$1(x^0)(3y)^7 = 2\,187y^7$. El coeficiente es 2 187.

5. Calcula el valor de b en: $(2x^3+b)^4 = 16x^{12}+96x^{11}+216x^{10}+216x^9+81x^8$

Sol. Para el segundo término se calculan las combinaciones $\binom{4}{1}$

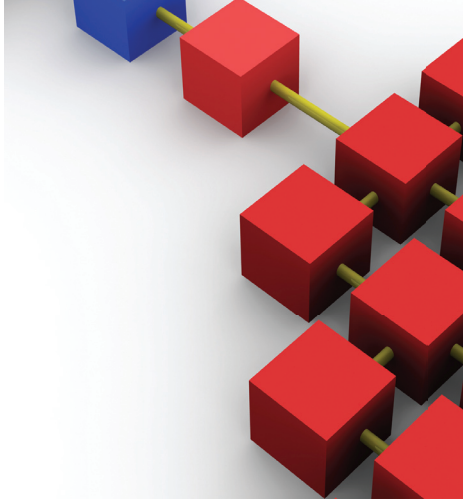
Entonces $\binom{4}{1}(+2x^3)^3(b)$ es el segundo término.

Igualando $\binom{4}{1}(+2x^3)^3(b) = 96x^{11}$

Resolviendo: $4(+8x^9)b = 96x^{11}$

$$b = \frac{96x^{11}}{+32x^9}$$

$$b = +3x^2$$



Capítulo 5

Estadística descriptiva

Introducción

La **estadística descriptiva** maneja los datos obtenidos para su ordenación y presentación, y hace resaltar ciertas características para que sean más objetivas y útiles.

La estadística descriptiva investiga los métodos y procedimientos, y establece reglas para que el manejo de los datos sea eficiente para que la información presentada resulte confiable, exprese en lenguaje sencillo los contenidos para que el mayor número de personas lo comprenda y así pueda establecer comparaciones y obtener conclusiones.

La **investigación estadística** es la operación que se refiere a la recopilación de información sobre una población o colectivo de individuos, medidas u objetos que tienen una característica común e incluye:

a) Señalamiento del elemento de la población que origina la información (unidad de investigación), la cual puede ser una industria, un hogar, una persona, etcétera. Sin embargo, en todo caso, la unidad debe ser en su definición medible y fácilmente identificable.

b) Citar qué se investiga, cómo se debe realizar, cuándo se llevará a cabo y el lugar de la investigación, que es el dónde.

c) La recolección de la información incluye ordenarla, filtrarla (eliminando posibles errores) y analizarla, aplicando los métodos y normas estadísticos.

d) La publicación de la información, ya sea para uso propio o ajeno.

Conceptos clave

Estadística descriptiva
Investigación estadística
Representación tabular
Cuadros cronológicos
Gráficos de líneas
Pictogramas
Gráficos de barras
Gráficos circulares

Presentación de la información

Una vez que se obtiene la información, resultado de una investigación estadística, que bien puede haberse efectuado, por ejemplo, en medicina, para estudiar el comportamiento de enfermos sujetos a un tratamiento específico; en educación, sobre los ensayos orientados a estudiar los cambios de actitud y aprendizaje de alumnos sometidos a ciertos procesos educativos; o bien, en la agricultura, dirigidos a medir el efecto de un insecticida bajo ciertas condiciones que varían bajo el control del investigador, es necesario escoger la forma de organizarla para su análisis o para su publicación. Estas formas de representación pueden ser en:

- Cuadros numéricos de información
- Gráficos y pictogramas

Cuadros numéricos de información

A. Representación tabular

Las líneas horizontales y las columnas verticales deben disponerse de manera que resalten los aspectos que se desean mostrar y las comparaciones que se quiere hacer notar.

Incluirá:

a) Título. Donde se indica el objeto del cuadro.

b) Columna principal. Lugar donde se anotan las categorías.

c) Encabezado de las columnas. Donde se explica el objeto de cada una de ellas.

d) Cuerpo. Lugar donde se pone la información.

e) Notas a pie de página. Ahí se aclaran algunas operaciones y se indica la fuente de la información.

Problemas:

1. El contador de una compañía industrial informa que durante el mes de marzo pasado el total de ventas fue de $11 745 420 y la nómina de pago del mes, por departamento, fue así:

Personal administrativo: $425 760

Personal de ventas y promoción: $528 750

Personal de producción: $2 765 450

Elabora un cuadro que señale:

a) Porcentaje de cada departamento con relación al total de la nómina.

b) Porcentaje de cada departamento con relación al total de ventas.

Solución:

Nómina de pago por departamento Mes de marzo Total de ventas en el mes $11 745 420			
Departamento	**Gastos mes**	**% nómina**	**% ventas**
Administración	425 760	11.45	3.62
Ventas	528 750	14.21	4.50
Producción	2 765 450	74.34	23.54
Totales	$3 719 960	100.00	31.66

Operaciones que hicimos para llenar el cuadro: calculamos por interpolación polar.

(Razones y proporciones):

Nómina:

$$3\,719\,960 : 100 :: 425\,760 : x$$

$$3\,719\,960x = 425\,760\,(100)$$

$$x = \frac{42\,576\,000}{3\,719\,960}$$

$$x = 11.45\%$$

$$3\,719\,960 : 100 :: 528\,750 : x$$
$$3\,719\,960x = 528\,750\,(100)$$
$$x = \frac{52\,875\,000}{3\,719\,960}$$
$$x = 14.21\%$$

$$3\,719\,960 : 100 :: 2\,765\,450 : x$$
$$3\,719\,960x = 2\,765\,450\,(100)$$
$$x = \frac{276\,545\,000}{3\,719\,960}$$
$$x = 74.34\%$$

Ventas:

$$11\,745\,420 : 100 :: 425\,760 : x$$
$$11\,745\,420x = 425\,760\,(100)$$
$$x = \frac{42\,576\,000}{11\,745\,420}$$
$$x = 3.62\%$$

$$11\,745\,420 : 100 :: 528\,750 : x$$
$$11\,745\,420x = 528\,750\,(100)$$
$$x = \frac{52\,875\,000}{11\,745\,420}$$
$$x = 4.50\%$$

$$11\,745\,420 : 100 :: 2\,765\,450 : x$$
$$11\,745\,420x = 2\,765\,450\,(100)$$
$$x = \frac{276\,545\,000}{11\,745\,420}$$
$$x = 23.54\%$$

2. Un representante de la Secretaría de Gobernación asistió a un sorteo organizado por una casa que vende material deportivo. En el evento se entregaron tres premios consistentes, cada uno, en un viaje para 2 personas a Rotterdam, Holanda. El representante entregó el siguiente reporte:

En la primera extracción, el premio se otorgó al boleto número 007950 y correspondía a Manuel López Galicia; en la segunda extracción, el premio se entregó al boleto 015162, correspondiente a María Roy Martínez; en la tercera extracción, el premio fue para el boleto número 008032, correspondiente a Yolanda Uribe May. Elabora el cuadro que incluya esta información.

Cuadro de ganadores promoción Deportes Parti Permiso de Gobernación con número S – 0322 – 2000 Sorteo realizado el día 20 de junio de 2000			
Número de extracción	**Número de folio**	**Nombre del ganador**	**Premio**
1	007950	Manuel López Galicia	Viaje a Rotterdam
2	015162	María Roy Martínez	Viaje a Rotterdam
3	008032	Yolanda Uribe May	Viaje a Rotterdam

B. Cuadros cronológicos

Se usan para expresar las variaciones cronológicas de población, producción, salarios, etcétera; el periodo que se cita en estos cuadros depende de lo que se desea comprar o mostrar.

Problema:

1. Elabora un cuadro cronológico de ganancias de una fábrica de piezas de motor en el quinquenio 1994-1998 que exprese:

a) Las variaciones de cada año en tanto por ciento con relación al año anterior.

b) Del año 1998 con relación al año 1994.

Si las ganancias en miles de pesos fueron:

1994 = 575

1995 = 644

1996 = 730.94

1997 = 672.47

1998 = 749.80

Solución:

Ganancias de la compañía en miles de pesos durante el quinquenio 1994-1998			
Año	**Ganancia**	**% Variación**	
		Base año anterior	**Base año 1994**
1994	575		
1995	644	12	
1996	730.94	13.5	
1997	672.47	–8	
1998	749.80	11.49	30.4

Operaciones

Con la interpolación polar:

$$575:100::644:x$$
$$575x=644\,(100)$$
$$x=\frac{64\,(100)}{575}=112\%$$

El 112 % significa que la ganancia de 1995 fue de 12% más de la obtenida en 1994 (que es el 100%).

Para las demás, razonamos en forma semejante.

$$644:100::730.94:x$$
$$644x=730.94\,(100)$$
$$x=\frac{73\,094}{644}=113.5\%$$
$$113.5-100=13.5\%$$

$$730.94:100::672.47:x$$
$$730.94x=672.47\,(100)$$
$$x=\frac{67\,247}{730.94}=92\%$$
$$92-100=-8\%$$

$$672.47:100::749.80:x$$
$$672.47x=749.80\,(100)$$
$$x=\frac{74\,980}{672.47}=111.49\%$$
$$111.49-100=11.49\%$$

$$575:100::749.80:x$$
$$575x=749.80\,(100)$$
$$x=\frac{74\,980}{575}=130.4\%$$
$$130.4-100=30.4\%$$

Gráficos y pictogramas

La forma de presentar esta información por medio de ideográficos dependerá del nivel cultural del auditorio a que va dirigido y del lugar de exposición (periódicos, revistas, televisión, escuelas). Los métodos más usuales son gráficos de líneas, pictogramas o pictógrafos, gráficos de barras y gráficos circulares.

A. Gráficos de líneas

Se usan para representar las distribuciones de frecuencias, tema que analizaremos más adelante.

Los gráficos son una representación estadística de gran utilidad para dar a conocer una idea global sobre un problema en que se aplican procedimientos estadísticos. Los datos que proporcionan son aproximados y por ello, se debe ser cuidadoso en su elaboración. Si en los gráficos se dibujan simultáneamente varios diagramas, la vista del usuario tendrá dificultad para identificarlos, aunque éstos se hayan diferenciado con colores o con diferentes tipos de trazo.

Además, la cantidad de información que proporciona un gráfico no es tan completa y extensa como la de un cuadro que tiene varias columnas que se leen por separado.

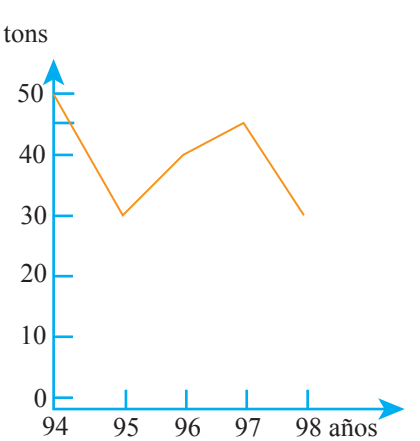
Al trazar un gráfico de líneas (diagramas lineales) debes tomar en cuenta los siguientes conceptos:

- La curva debe ser más gruesa que las coordenadas para que resalte.
- La unidad de medida que se utilice debe destacarse claramente, no necesariamente de un centímetro.
- La longitud se seleccionará de modo que la gráfica resulte balanceada.
- En las notas a pie de página se deben citar los conceptos aclaratorios de la curva.
- Siempre debe colocarse el cero de la escala vertical.
- De ser posible, hay que citar la fuente de la información.
- Se localizan, por las coordenadas correspondientes, los puntos de interés y se unen por segmentos de rectas, formándose así una poligonal, que es el diagrama de la serie cronológica.
- Es importante tener cuidado con la escala de los ejes porque es posible manejarlos en forma engañosa, como se puede apreciar en el siguiente problema.

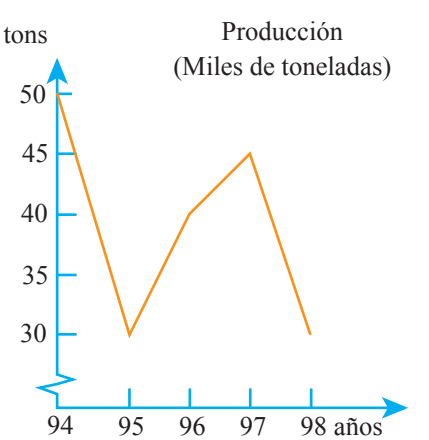
Problemas

1. Una compañía industrial trata de vender acciones y su departamento de contabilidad presenta dos gráficas sobre su producción en el periodo de 1994-1998. Decide cuál de las dos gráficas presenta los datos con mayor veracidad.

Solución:



Esta gráfica es la más veraz



Las dos gráficas presentan hechos reales, pero en los diagramas se crearon dos imágenes diferentes para un mismo suceso estadístico alterando los valores del eje vertical y la unidad de medida en la horizontal.

2. Consulta un periódico de circulación nacional y observa el índice UV del día que decidas. El índice UV se refiere al daño que los rayos ultravioleta pueden hacer a un ser humano.

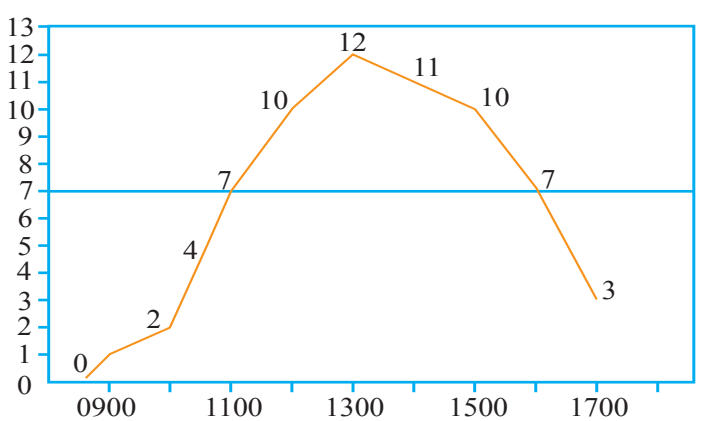
Cuando el índice UV está por encima de 9, los rayos UV son extremadamente fuertes y la piel sufrirá quemaduras en menos de 15 minutos. Los periodos de quemadura de piel por exposición al sol están calculados con base en una piel clara no bronceada; el lapso de tiempo sería un poco más prolongado para aquellos con piel más oscura.

Solución:

Índice UV de rayos ultravioleta que corresponden a un día del mes de junio de 2000

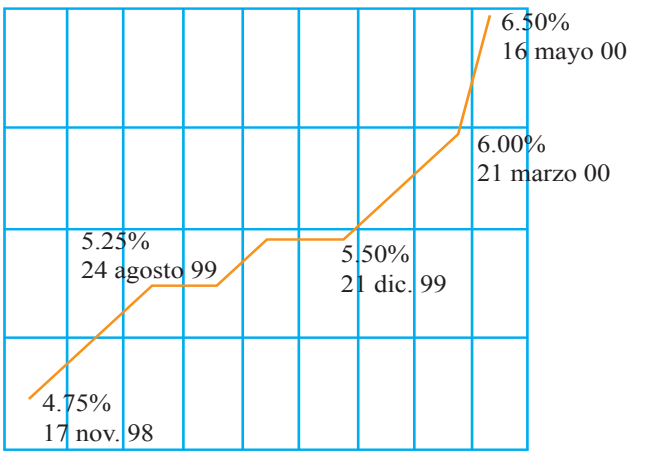
Índice	Exposición al Sol	Calificación
Más de 9	Menos de 15 min	Extremo 50
De 7 – 9	20 min	Alto
De 4 – 7	20 min	Moderado
De 0 – 4	Más de una hora	Bajo

Gráfica de UV



La raya horizontal señala las horas de más intensidad del sol.
Fuente: Periódico *Reforma*. Junio 2000.

3. Se cita a continuación una gráfica que señala la tendencia a la alza de las tasas de interés internacionales. ¿Qué concluyes?



Fuente: Departamento de análisis, periódico *Reforma* con datos federales. Junio 2000.

Ahí permanecerá, excepto que en fecha próxima sea necesario encarecer el dinero para bajar el consumo y así evitar presiones inflacionarias.

B. Pictogramas

Un pictograma es la representación de datos estadísticos con símbolos que, por su forma, sugieren la naturaleza del dato. Se utiliza para expresar comparaciones que atraigan la atención general, cualquiera que sea el nivel cultural del lector. Su representación no sirve para análisis estadísticos y únicamente permite obtener conclusiones válidas muy generales.

Al hacer la representación con un pictograma o pictógrafo se deben utilizar figuras del mismo tamaño; las aproximaciones se hacen con fracción de la figura, mitad y hasta cuartos, y la cantidad que representa cada figura se indica con claridad en el encabezado.

Problema

1. La información oficial preliminar del INEGI señala que el total de la población que habita la República Mexicana es de 97.4 millones, de los cuales 47.4 millones son hombres y 50 millones son mujeres; de todos éstos, 24.64 millones es población rural, 72.76 millones urbana y dentro de la urbana, 17.79 millones corresponde a la zona urbana del Valle de México.

El reporte establece además que la tasa de crecimiento anual fue: en la década de 1980-1990 de 2.4%; en el quinquenio 1990-1995 el 2.1% y de 1995-2000 disminuyó a 1.6%; por la tasa de crecimiento ocupamos el sexto lugar en el mundo.

En l980 había 66.8 millones de habitantes, en 1990 había 81.3 millones y para el 2000, 97.4 millones, lo que hacía que la República Mexicana ocupara el onceavo lugar más poblado del mundo.

El crecimiento absoluto por estados es (en millones de habitantes) el siguiente:

- Estado de México: 3.27
- Puebla: 9.4
- Nuevo León: 0.73
- Jalisco: 1.02
- Baja California: 0.83
- Otros estados: 9.31

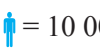
Los estados más poblados, en millones de habitantes, son:

- Estado de México: 13.08
- Veracruz: 6.90
- Puebla: 5.07
- Distrito Federal: 8.59
- Jalisco: 6.32

Representa gráficamente esta información.

Solución:

a) Población en la República Mexicana: 97.4 millones de habitantes.

 = 10 000 000 habitantes

Hombres 47.4

Mujeres 50

Rural 24.64

Urbana 72.76; de éstos en el D.F. y zona conurbada son 17.79

b) Distribuidos así:

Aumentó la población de 1990 al 2000, en:		
1980	éramos	66.8
1990		81.3
2000		97.4

Del aumento de 97.4 – 81.3 = 16.1 se repartieron así:	
Estado de México	3.27
Jalisco	1.02
Puebla	0.94
Baja California	0.83
Nuevo León	0.73
Otros estados	9.31

Estados más poblados (millones de habitantes)	
Estado de México	13.08
Distrito Federal	8.59
Veracruz	6.90
Jalisco	6.32
Puebla	5.07

Crecimiento: disminuyó	
1980-1990	2.4%
1990-1995	2.1%
1995-2000	1.6%

Conclusión:

Con base en los nacimientos entre 1980-1990 de 2.4%, en la actualidad la demanda de estos jóvenes es alta en las escuelas de enseñanza media superior y superior; en cambio, por los nacidos entre 1990-1995 su asistencia en primaria es baja y lo será menos para los nacidos entre 1995-2000, apenas grupos de 20 a 25 alumnos.

Gráficos de barras

Los gráficos de barras proporcionan más información y permiten una apreciación estadística más clara que los pictogramas. Se utilizan para representar datos nominales y variables cardinales.

Para su elaboración, debes tomar en cuenta lo siguiente:

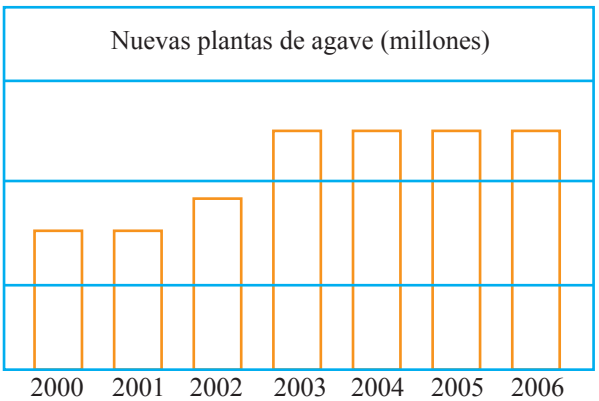
- En el gráfico se debe evitar que las barras resulten muy anchas o excesivamente altas.
- Debes dejar un espacio entre las barras, que no sea inferior a la mitad del ancho de ellas.
- Si el gráfico incluye muchas barras, es mejor sustituirlo con un diagrama lineal.

Problemas

1. Una fuente de trabajo y de divisas extranjeras al país es la venta de tequila en los mercados de Japón, Alemania, Estados Unidos y otros países. Sin embargo, la demanda aumenta y el agave escasea cada vez más; por ello, los industriales del ramo han decidido plantar en los próximos siete años 263 millones de hijuelos de agave para evitar la escasez. Así, en el presente año y el próximo se plantarán 35 millones en cada uno; en el 2002, 37 millones y en cada uno de los restantes 39 millones.

Expresa esta solución con un gráfico de barras.

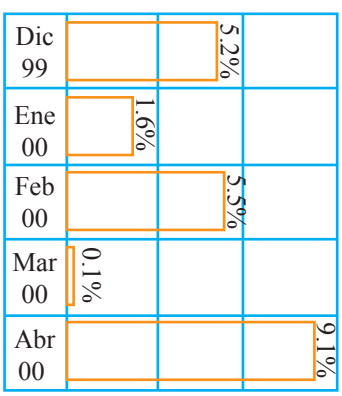
Solución:



Fuente: Consejo regulador del tequila
Periódico *Reforma*. Junio 2000.

Estas barras también se pueden disponer en forma horizontal.

2. El siguiente gráfico de barras expresa las ventas en tiendas de autoservicio y departamentales en el mes de diciembre de 1999 y los de enero a abril de 2000. ¿Qué puedes concluir?



Conclusión:

Hay mucho dinero circulante que no corresponde a nuestra capacidad de producción.

Cuando el consumo aumenta y las personas empezamos a gastar en cosas innecesarias, las autoridades económicas, con el fin de evitar presiones inflacionarias, reducen el circulante con un "corto".

Gráficos circulares

Se usan para presentaciones gráficas de distribuciones porcentuales y si se quieren utilizar para secuencias cronológicas, es necesario dibujar círculos iguales, uno por cada año, señalando en cada uno la correspondiente distribución porcentual.

El círculo de 360° corresponde a un área de 100%, un sector representa un tanto por ciento equivalente a la razón entre el ángulo que forman los radios que limitan el sector y los 360°, que son el total de grados de la circunferencia. Analiza los siguientes datos:

$\frac{180}{360} = 0.5$, o sea el 50%

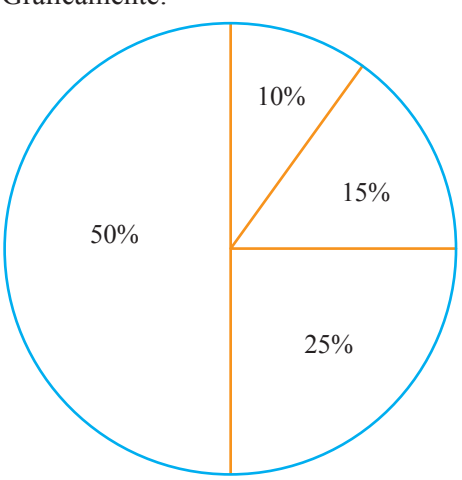
$\frac{90}{360} = 0.25$, o sea el 25%

$\frac{54}{360} = 0.15$, o sea el 15%

$\frac{36}{360} = 0.10$, o sea el 10%

Se tomaron 2 cifras decimales.

Gráficamente:



Problema

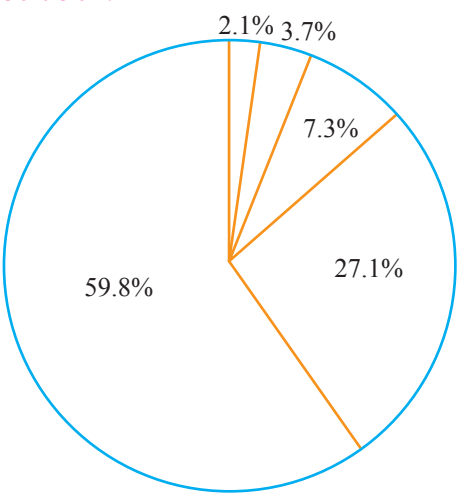
1. El gas natural es uno de los principales insumos para la generación de electricidad a través de las termoeléctricas. Tiene además usos para la industria y los hogares.

La Secretaría de Energía y PEMEX señalan que la capacidad instalada del sector paraestatal por generación de energía en el año de 1999 fue de 35 675.1 megawatts, generados así:

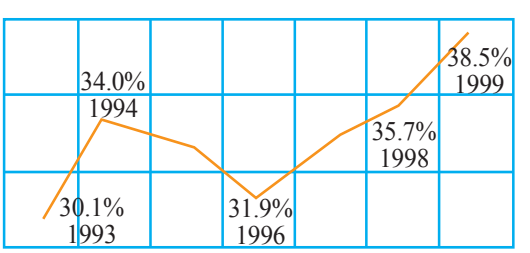
Termoeléctrica	21 351.1, o sea el 59.8%
Hidroeléctrica	9 662.8, o sea el 27.1%
Carboeléctrica	2 600, o sea el 7.3%
Nucleoeléctrica	1 309, o sea el 3.7%
Geotermoeléctrica y eoloeléctrica	752.1, o sea el 2.1%

De estas fuentes, la carboeléctrica resulta la más contaminante. Representa esta información en un gráfico circular.

Solución:



La industria eléctrica demanda mucho gas natural.



Fuente: Secretaría de Energía y PEMEX. Junio 2000.

A mayor industrialización, que así se espera con los nuevos tratados económicos, mayor número de empleos, mayor demanda de energía eléctrica y encarecimiento del gas natural, industrial y doméstico.

Procura que en tu casa, de ser posible, se instale un aparato que capte la energía solar; en países como Japón, Israel y Estados Unidos lo usan con éxito y disponen de pocos meses en que hay sol.

Hay estados como el de Morelos, Zacatecas y otros en los que más de 90% de los días del año son soleados.

Ejercicios de repaso

I. Considerando el siguiente cuadro, contesta las preguntas 1-6.

Población total según sexo, 1950 a 2005			
Año	Total	Hombres	Mujeres
1950	25 791 017	12 696 935	13 094 082
1960	34 923 129	17 415 320	17 507 809
1970	48 225 238	24 065 614	24 159 624
1990	81 249 645	39 893 969	41 355 676
1995	91 158 290	44 900 499	46 257 791
2000	97 483 412	47 592 253	49 891 159
2005	103 263 388	50 249 955	53 013 433

Fuente:
INEGI Censos de población y vivienda, 1950 y 2000.
INEGI Censos de población y vivienda, 1995 y 2005.

1. ¿Qué tipo de gráfico se utilizó? _____

2. Escribe el título, columna principal, encabezado de las columnas y notas a pie de página.

_____

_____

3. ¿Cuál fue el total de población en el 2005? _____

4. En el 2005, ¿cuál era el porcentaje de hombres y mujeres respectivamente?

_____

5. Con los datos del cuadro, construye un cuadro cronológico con los datos de hombres y mujeres en porcentaje.

6. De 1950 a la fecha, ¿se ha conservado el porcentaje de mayoría de mujeres?

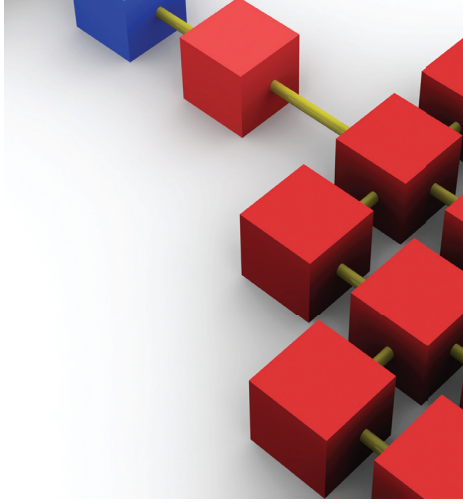
II. En una escuela secundaria se aplicó un examen diagnóstico en la materia de matemáticas a los alumnos de tercer grado. En el grupo 3A se aplicó a 36 alumnos y 20 lo aprobaron; en el 3B se aplicó a 34 y lo aprobaron 18; en el 3C lo presentaron 35 alumnos y lo aprobaron 19 y por último, en el 3D lo presentaron 29 alumnos y lo aprobaron 21. Cada grupo consta de 36 alumnos.

Con esta información responde las preguntas 7-9.

7. Construye una representación tabular que indique el número de alumnos que presentó el examen, número de alumnos que lo aprobó, porcentaje de alumnos que lo aprobó respecto al total de alumnos, porcentaje de alumnos que lo aprobó respecto al número de alumnos que lo presentó.

8. ¿Cuál es el grupo que tuvo mayor porcentaje de aprobación? _____

9. ¿Cuál es el grupo que tuvo mayor porcentaje de reprobación? _____



Capítulo 6

Probabilidad

Introducción

El concepto intuitivo de la **predicción** por medio del cual una persona toma decisiones sin la certeza de que ocurran todos sus supuestos es la base de un estudio sistemático denominado **probabilidad**, que permite incrementar el grado de confianza para decidir.

A pesar de que el conocimiento de la probabilidad nos permite saber qué creemos que va a suceder, no nos ayuda a saber qué va a suceder de manera precisa.

Los diferentes tipos de probabilidades a los que nos referimos se basan en:

I. La frecuencia relativa.

II. Los sucesos compuestos. Probabilidad axiomática.

III. La probabilidad condicional.

Conceptos clave

- Predicción
- Probabilidad
- Experimento aleatorio
- Frecuencia relativa
- Probabilidad empírica
- Fórmula básica de la probabilidad
- Modelo definido de regularidad
- Cuaternas ordenadas
- Composición de sucesos

Probabilidad como frecuencia relativa

Consideraciones generales

El concepto de la probabilidad con base en la frecuencia relativa es la más sencilla y la que más utilizan las personas para tomar una decisión.

Un **experimento aleatorio** es aquel que al realizarlo no sabemos exactamente qué resultado tendrá pero sí sabemos cuántas cosas diferentes pueden suceder.

Por ejemplo, si en una bolsa tenemos pelotas amarillas, rojas y azules, y luego tomamos una sin ver, no sabremos de qué color saldrá, pero sabremos con certeza que la pelota será amarilla, roja o azul.

Al repetir el experiento un número considerable de veces, una fracción de este número produce un suceso o evento E_1 (que salga pelota amarilla), otra el suceso E_2 (pelota azul), otra el suceso E_3 (pelota roja) y así sucesivamente si tuviéramos pelotas de otros colores.

Aunque las palabras suceso o evento son sinónimos, en este libro usaremos la palabra suceso y tú, la que proponga el profesor.

Si un experimento se repite muchas veces, digamos n y si el suceso E_1 se observa h veces, entonces la probabilidad S del suceso E es el cociente de la razón $\frac{h}{n}$.

$$\text{Probabilidad } S = \frac{\text{Número de veces que el suceso } E_1 \text{ ocurrió}}{\text{Total de sucesos realizados}} = \frac{h}{n}$$

La experiencia justifica esta igualdad porque a medida que n crece, la frecuencia relativa se aproxima más a la probabilidad matemática.

Este concepto se utiliza para definir la razón citada como **probabilidad empírica**, término que algunos autores conocen como **fórmula básica de la probabilidad**. Esta fórmula es muy utilizada en la interpretación de datos estadísticos y de seguros.

Problema 1

¿Cuál es la probabilidad de que una persona de 25 años de edad llegue a sobrevivir hasta los 40 si de acuerdo con una tabla de mortalidad, de cada 93 745 personas de 25 años de edad sólo 87 426 llegan a los 40 años.

Solución:

Como $h = 87\,426$

$n = 93\,745$

$$\text{Probabilidad } S = \frac{\text{Personas que llegan a los 40 años}}{\text{Total de personas de 25 años}} = \frac{87\,426}{93\,745}$$

$$= 0.9325 \text{ (se tomaron 4 cifras decimales)}$$

Probabilidad expresada en tanto por ciento

Otra forma de expresar el resultado obtenido de la frecuencia relativa es en tanto por ciento. Así, podemos expresar como porcentaje el resultado del ejemplo anterior si lo multiplicamos por 100.

$$\text{Probabilidad } S = \frac{87\,426}{93\,475}(100) = 0.9325(100) = 93.25\%$$

Problema 2

En una caja hay 25 tornillos en buen estado y 80 defectuosos. ¿Cuál es la probabilidad de sacar de la caja un tornillo en buen estado?

Solución:

Como $h = 25$

$n = 80 + 25$

$$\text{Probabilidad } S = \frac{\text{Número de tornillos en buen estado}}{\text{Total de tornillos en la caja}} = \frac{25}{25+80} = \frac{25}{105}$$

$$= 0.2380 \text{ (Se tomaron 4 cifras decimales)}$$

Probabilidad en porcentaje = 0.2380 (100) = 23.80%

Problema 3

De cada 1 000 personas a las que se les practica una revisión médica, 35 tienen problemas de la vista. ¿Cuál es la probabilidad de que una persona examinada tenga alguna enfermedad en los ojos?

Solución:

Como $h = 35$

$n = 1\,000$

$$\text{Probabilidad } S = \frac{\text{Número de personas con problemas de la vista}}{\text{Total de personas examinadas}} = \frac{35}{1\,000} = \frac{7}{200}$$

$$= 0.035$$

Probabilidad en porcentaje = 0.035 (100) = 3.5%

Problema 4

En una caja hay 75 canicas azules y 225 rojas. ¿Cuál es la probabilidad de obtener al azar una canica azul?

Solución:

Como $h = 75$

$n = 75 + 225$

$$\text{Probabilidad } S = \frac{\text{Número de canicas azules}}{\text{Total de canicas en la caja}} = \frac{75}{75+225} = \frac{75}{300}$$

$$= 0.25$$

Probabilidad en porcentaje = 0.25 (100) = 25%

Propiedades de la frecuencia relativa

1) Para cada suceso A:

$0 \leq P(A) \leq 1$

La probabilidad de un suceso está entre 0 y 1.

2) $P(A) = 1$ Si y sólo si A *ocurre* en las n repeticiones.

3) $P(A) = 0$ Si y sólo si A *nunca ocurre* en las n repeticiones.

Probabilidad de que ocurra o no un suceso

Un suceso A es seguro si se observa n veces de las n ocasiones en las que se realiza el experimento para cualquier n.

Entonces, la probabilidad de $P(A)$ es igual a $\frac{n}{n} = 1$. Se dice que A es un *suceso seguro*.

Si el suceso contiene la totalidad de resultados posibles del experimento, entonces el conjunto asociado es el universo y se dice que es un suceso seguro. La probabilidad de un suceso seguro es uno.

La probabilidad de un suceso es p y la probabilidad de que no ocurra es q. Esto se expresa así:

$$q = 1 - p$$

Si ningún resultado es favorable, el suceso se define a través del conjunto vacío, se trata de un *suceso imposible*. La probabilidad de un *suceso imposible es cero*.

Consideremos además:

- Si sólo un resultado es favorable, el suceso se asocia a un conjunto de un solo elemento. Se dice entonces que el *suceso es simple*.
- Si un grupo de resultados es favorable, el conjunto asociado tiene varios elementos. Se dice entonces que es un *suceso compuesto* porque está formado por la unión de sucesos simples.
- *Suceso contrario* es el que se cumple cuando no se cumple A; se denota con A', que equivale al *conjunto complemento*.

Problema 5

En una caja hay 10 boletos marcados con un dígito: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Si se obtiene al azar un boleto de la caja, identifica los sucesos siguientes:

A: Sacar un número mayor que 10.

B: Sacar un número primo mayor que 2 y menor que 5.

C: Sacar el número 2.

D: Sacar un número que sea el doble del 3.

E: Sacar un número que sea la mitad del 2.

F: Sacar un número que sea un divisor del 6.

G: Sacar un número natural menor que 10.

¿Cuál es el conjunto universal U de este experimento?

Solución:

$A = \varnothing$ suceso imposible, ningún resultado lo favorece.

$B = \{3\}$ suceso simple

$C = \{2\}$ suceso simple

$D = \{6\}$ suceso simple

$E = \{1\}$ suceso simple

$F = \{1, 2, 3, 6\}$ evento compuesto formado por la unión de sucesos simples. Se expresa así: $F = E \cup C \cup B \cup D$

$G = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ suceso seguro, siempre ocurre

$U = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, de donde $G = U$

Datos de un problema

Se obtienen con experimentación controlada o por observación de los sucesos incontrolados de la naturaleza.

Población

Es el conjunto de todos los sucesos susceptibles de aparecer en un problema y que interesan a la persona que realiza el estudio.

Experimento aleatorio

Un experimento aleatorio que se identifica con la letra E es un proceso que al repetirse varias veces permite la observación de todo lo que puede ocurrir bajo ciertas circunstancias.

Los resultados de un experimento aleatorio dan lugar a un espacio muestral S.

Problema 6

Consideremos los siguientes experimentos aleatorios:

- E_1: Se lanza una moneda al aire y se observa que cae águila o sol.
- E_2: Se analiza la calidad de los objetos producidos por una fábrica en 8 horas. Se observa que unos son de buena calidad y otros son defectuosos.
- E_3: Se toman varios fusibles de automóvil y se registra el tiempo que tardan en quemarse.
- E_4: En una caja se guardan 50 pelotas de colores: azules, blancas y rojas. Se toma al azar una bola y se registra su color.

Estos experimentos tienen en común los siguientes aspectos:

a) Podemos observar muchas veces cada experimento sin cambiar en lo esencial las condiciones.

b) Podemos describir el espacio muestral S de todos los resultados posibles del experimento.

c) A medida que repetimos el experimento, los resultados de cada uno parecen u ocurren en forma caprichosa.

Sin embargo, como el experimento se repite gran número de veces, finalmente aparece un **modelo definido de regularidad**. Esta regularidad de resultados permite establecer un modelo matemático con el que podemos analizar el experimento.

Muestra

Es un subconjunto de mediciones seleccionadas de la población que fundamenta un problema.

Problema 7

Experimento aleatorio:

Analiza la calidad de los objetos que se producen en una fábrica en 8 horas de trabajo. La población está dada por todos los objetos producidos que sabemos pueden ser o no defectuosos.

Dado que no sería razonable ni posible revisar cada uno de los productos, debemos trabajar con una muestra. Es decir, debemos tomar cierto número de objetos que se produjeron en un lapso de 8 horas. Así, contamos cuántos objetos se produjeron y cuántos de ellos son defectuosos. Con estos datos calculamos la probabilidad del número

de objetos defectuosos en un periodo de 8 horas y así, dependiendo del resultado de la probabilidad se puede decidir si la producción es o no de calidad.

Problema 8

Experimento aleatorio:

Se lanzan dos dados, uno rojo y otro azul. Señala cuántos puntos muestrales correspondan a este experimento y expresa el espacio muestral.

Solución:

Como cada dado tiene 6 caras y cada resultado se forma de dos colores, tenemos que el número de puntos muestrales es:

N = (6 del dado rojo) (6 del dado azul) = 36 puntos muestrales.

El espacio muestral se puede definir por **parejas ordenadas**, donde el primer elemento es el resultado del dado rojo y el segundo es el resultado del dado azul. Así:

$$S = \{(1,1)\ (1,2)\ (1,3)\ (1,4)\ (1,5)\ (1,6)$$
$$(2,1)\ (2,2)\ (2,3)\ (2,4)\ (2,5)\ (2,6)$$
$$(3,1)\ (3,2)\ (3,3)\ (3,4)\ (3,5)\ (3,6)$$
$$(4,1)\ (4,2)\ (4,3)\ (4,4)\ (4,5)\ (4,6)$$
$$(5,1)\ (5,2)\ (5,3)\ (5,4)\ (5,5)\ (5,6)$$
$$(6,1)\ (6,2)\ (6,3)\ (6,4)\ (6,5)\ (6,6)\}$$

Problema 9

Experimento aleatorio:

Se lanza un dado sobre la mesa. Señala los puntos muestrales de los sucesos siguientes:

A: Salga un número 5.

B: Salga un número impar.

C: Salga un número menor que 3.

D: Suceda simultáneamente A y B.

E: Suceda simultáneamente A y C.

Solución:

$A = \{5\}$

$B = \{1, 3, 5\}$

$C = \{1, 2\}$

$D = \{5\}$

$E = \varnothing$

Problema 10

Experimento aleatorio:

Se lanzan 4 monedas distintas de $1.00, $5.00, $10.00 y de $20.00; al caer, señala cuántos puntos muestrales son y describe el espacio muestral.

Solución:

Como cada moneda puede caer en dos formas distintas, tenemos:

$N = (2)\ (2)\ (2)\ (2) = 16$ puntos muestrales.

El espacio muestral se puede definir como cuaternas ordenadas, donde el primer elemento es el resultado de la moneda de un peso; el segundo es de $5.00, el tercero de $10.00 y el cuarto de $20.00. Así:

$$S = \{(a, a, a, a)$$
$$(s, a, a, a)\ (a, s, a, a)\ (a, a, s, a)\ (a, a, a, s)$$
$$(s, s, a, a)\ (s, a, s, a)\ (s, a, a, s)\ (a, s, s, a)\ (a, s, a, s)\ (a, a, s, s)$$
$$(s, s, s, a)\ (s, s, a, s)\ (s, a, s, s)\ (a, s, s, s)$$
$$(s, s, s, s)\}$$

Observa que para encontrar los elementos de S primero se considera que al lanzar las monedas no salga sol, luego se escriben todos los sucesos posibles en los que una de las monedas sale sol, después en los que salen dos soles; se sigue con todos los sucesos en los que salen tres soles y finalmente, se registra cuando salen cuatro soles.

Tipos de sucesos

En función de la relación de probabilidad que se puede establecer entre los sucesos, éstos se clasifican en:

a) Mutuamente excluyentes o disjuntos

Son sucesos en los que en un mismo experimento aleatorio no es posible que ocurran simultáneamente. La intersección de los conjuntos que los representan es el conjunto vacío.

En sucesos mutuamente excluyentes se tiene que la ocurrencia de uno de ellos elimina automáticamente la posibilidad de que ocurra el otro.

b) No excluyentes entre sí

Son aquellos sucesos que se presentan en un mismo experimento aleatorio y en los que la posibilidad de que ocurra uno de ellos no impide que el otro suceso ocurra; es decir, pueden ocurrir conjuntamente.

La intersección de los conjuntos que los representan es diferente al conjunto vacío.

Problema 11

Experimento aleatorio:

Lanzar un dado.

Consideremos los sucesos siguientes:

A: Salga el número 4.

B: Salga un número primo.

C: Salga un número múltiplo de 2.

Solución:

$A \cap B = \varnothing$ Los sucesos A y B *son* mutuamente excluyentes.

$A \cap C \neq \varnothing$ Los sucesos A y C *no son* mutuamente excluyentes.

Problema 12

Experimento aleatorio:

Se analiza en un momento dado el estado de salud de los habitantes de una comunidad.

Consideremos los sucesos siguientes:

A: La persona es diabética.

B: La persona está sana.

C: La persona tiene un problema de salud permanente.

D: La persona tiene gripa.

E: La persona es hipertensa.

Solución:

$A \cap B = \varnothing$ Los sucesos A y B *son* mutuamente excluyentes puesto que una persona sana no puede ser diabética y si es diabética no está sana.

$C \cap E \neq \varnothing$ Los sucesos C y E *no son* mutuamente excluyentes porque en el conjunto de personas que tienen un problema de salud permanente están los que sufren de hipertensión; a la vez, un hipertenso es una persona con un problema de salud permanente.

$B \cap C = \varnothing$ Los sucesos B y C *son* mutuamente excluyentes porque una persona no puede considerarse a la vez sana y tener un problema de salud permanente.

$C \cap D \neq \varnothing$ Los sucesos C y D *no son* mutuamente excluyentes porque existe la posibilidad de que una persona con un problema de salud permanente esté enferma de gripa.

Problema 13

Experimento aleatorio:

El sorteo de la lotería.

Consideremos los sucesos siguientes:

A: Sacar reintegro.

B: Sacar el premio mayor.

Solución:

$A \cap B = \varnothing$ Los sucesos A y B *son* mutuamente excluyentes porque una persona que gane el premio mayor no puede tener reintegro, ni viceversa.

Problema 14

Experimento aleatorio:

Se observa la escolaridad de las personas de 20 a 60 años de edad de una comunidad.

Consideremos los sucesos siguientes:

A: Una persona tiene menos de 40 años.

B: La persona es ingeniero.

C: La persona es analfabeta.

D: La persona tiene 40 años o más.

Solución:

$A \cap B \neq \varnothing$ Los sucesos A y B *no son* mutuamente excluyentes porque es posible que una persona entre 20 y 40 años sea ingeniero, a la vez que un ingeniero puede tener menos de 40 años y más de 20.

$B \cap D \neq \varnothing$ Los sucesos B y D *no son* mutuamente excluyentes ya que un ingeniero puede tener más de 40 años y entre las personas de 40 a 60 años puede haber ingenieros.

$B \cap C = \varnothing$ Los sucesos B y C *son* mutuamente excluyentes puesto que una persona analfabeta no puede ser ingeniero y un ingeniero no es analfabeta.

$A \cap D = \varnothing$ Los sucesos A y D *son* mutuamente excluyentes porque una persona no puede tener menos de 40 años o más de 40 años en un mismo momento.

Probabilidad con base en los sucesos compuestos. Probabilidad axiomática

Consideraciones generales

Los problemas de probabilidad de mayor interés en situaciones prácticas son sucesos compuestos que para su solución requieren la enumeración de muchos puntos muestrales, procedimiento que resulta lento y cansado; de ahí que haya un segundo procedimiento denominado **composición de sucesos**.

La composición se forma con dos o más sucesos y se realiza con la unión o intersección de conjuntos o bien, con la combinación de ambos, y se basa en la clasificación de los sucesos, las relaciones entre ellos y tres leyes: la aditiva, la multiplicativa y la de sustracción.

Como nos estamos refiriendo a la probabilidad axiomática, es conveniente recordar que un axioma es una proposición matemática evidente por sí misma que no requiere demostración.

Unión de conjuntos

Dados dos conjuntos A y B se define el suceso $A \cup B$ como aquel que se cumple cuando se verifica A o B; tiene las mismas propiedades que la unión de conjuntos.

$A \cup S = S$, donde S es el espacio muestra

$A \cup \varnothing = A$

$A \cup A' = S$

Si $B \subset A$ entonces $A \cup B = A$

Intersección de conjuntos

Dados dos conjuntos A y B se define el suceso $A \cap B$ como el que se cumple cuando tienen lugar A y B a la vez. Tiene las mismas propiedades que la intersección de conjuntos.

$A \cap S = A$, donde S es el espacio muestra

$A \cap \varnothing = \varnothing$

$A \cap A' = \varnothing$

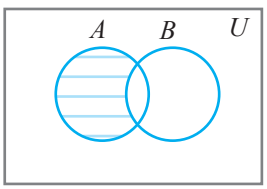
Si $B \subset A$ entonces $A \cap B = B$

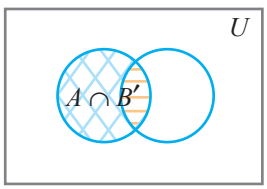
Todas las demás propiedades de la unión e intersección de conjuntos se verifican para la unión e intersección de sucesos.

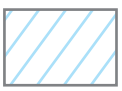
Diferencia de sucesos

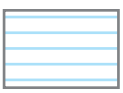
Dados dos sucesos A y B, se define el suceso $A - B$ como aquel que se cumple cuando se cumple A pero no B.

Se expresa $A - B = A \cap B'$

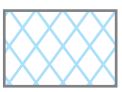








$A - B$



$A \cap B'$

La probabilidad de que un suceso u otro ocurra se calcula con las relaciones siguientes:

$P\,(A \text{ o } B) = P(A) + P(B)$

$P\,(A \text{ o } B) = P(A) + P(B) - (A \cap B)$

Para saber cuál de las dos relaciones debemos aplicar, es necesario revisar si los sucesos son o no mutuamente excluyentes.

a) Cuando dos sucesos son mutuamente excluyentes se tiene $A \cap B = \varnothing$; se utiliza entonces la primera relación.

La probabilidad de que A o B ocurran indistintamente es igual a la suma de sus probabilidades individuales.

b) Cuando los sucesos A y B no son mutuamente excluyentes se tiene $A \cap B \neq \varnothing$; se utiliza entonces la segunda relación.

Restar $P(A \cap B)$ tiene como función rectificar el doble conteo que se lleva a cabo cuando se suman $P(A)$ y $P(B)$.

Nota: Algunos autores expresan el suceso (A o B) como ($A \cup B$). Te recomendamos que tú lo hagas como lo indique tu profesor.

Problema 15

Para participar en la rifa de un reloj, los alumnos de primer año compraron 18 boletos y los de segundo grado 12 boletos. ¿Cuál es la probabilidad de que un alumno de primero o segundo gane la rifa? Se imprimieron 50 boletos.

Solución:

A: Un alumno de primer grado gana el premio.

B: Un alumno de segundo grado gana el premio.

El suceso que nos interesa es $E = A$ o B, los sucesos A y B son mutuamente excluyentes, es decir, $A \cap B = \varnothing$.

Entonces:

$$P(A \circ B) = P(A) + P(B) = \frac{18}{50} + \frac{12}{50} = \frac{30}{50} = \frac{3}{5} = 0.6$$

$$= 60\%$$

Problema 16

La tabla siguiente muestra el nivel de estudios de los profesores de una escuela.

	Escuela Nacional de Maestros	Escuela Normal Superior	Escuela normal privada	Especialización en la Universidad Pedagógica Nacional	Hombre	Mujer
M1	X			X		X
M2	X	X		X		X
M3		X	X			X
M4	X					X
M5	X			X		X
M6	X	X		X		X
M7			X	X		X
M8	X					X
H1		X	X	X	X	
H2	X			X	X	
H3	X				X	
H4	X	X		X	X	
H5	X				X	
H6			X	X	X	
H7	X	X		X	X	
H8			X		X	
H9	X			X	X	
H10			X	X	X	
18	12	6	6	12	10	8

Pregunta 1

¿Cuál es la probabilidad de que a un alumno le toque un profesor egresado de la Escuela Nacional de Maestros o que tenga una especialización en la Universidad Pedagógica Nacional?

Solución:

A: Profesor egresado de la Escuela Nacional de Maestros

B: Profesor egresado de la Universidad Pedagógica Nacional

$A \cap B \neq \varnothing$ Dado que hay docentes que son egresados de ambas instituciones. Los sucesos no son mutuamente excluyentes.

Entonces:

$$P(A \text{ o } B) = P(A) + P(B) - P(A \cap B)$$

$$= \frac{12}{18} + \frac{12}{18} - \frac{8}{18} = \frac{16}{18} = 0.8888$$

$$= 88.88\%$$

Nota: Explicación de cómo se obtiene la expresión $\frac{8}{18}$, en especial el valor del numerador 8. Analizamos la tabla citada al inicio del problema con relación a qué profesores, hombres o mujeres, cumplen la condición de ser egresados de la Escuela Nacional de Maestros y tienen una especialidad en la Universidad Pedagógica Nacional.

Observamos el primer renglón de la tabla (M1) y vemos que la primera persona asistió a las dos instituciones mencionadas; así sucesivamente, los renglones M2, M5, M6, H2, H4, H7, H9 cuya suma es de 8 personas, que es el valor obtenido.

Pregunta 2

¿Cuál es la probabilidad de que el profesor haya estudiado en la Escuela Normal Superior o sea egresado de una escuela normal privada?

Solución:

C: Profesor egresado de la Escuela Normal Superior

D: Profesor egresado de una escuela normal privada

$C \cap D \neq \varnothing$ Puesto que hay docentes que son egresados de una o de otra institución. Los sucesos no son mutuamente excluyentes.

Entonces:

$$P(C \text{ o } D) = P(C) + P(D) - P(C \cap D)$$

$$= \frac{6}{18} + \frac{6}{18} - \frac{2}{18} = \frac{10}{18} = \frac{5}{9} = 0.5555 = 55.55\%$$

El numerador de la expresión $\frac{2}{18}$ lo calculamos como en el caso anterior, analizando la tabla sobre los profesores, hombres y mujeres, que cumplen la condición de ser egresados de la Escuela Normal Superior y de una escuela normal particular. De acuerdo con la tabla, vimos que una persona que asistió a una o a otra institución es la del renglón M3 y la otra en H1, por lo que el valor obtenido es 2.

Pregunta 3

¿Cuál es la probabilidad de que el profesor sea egresado de la Escuela Normal Superior o de la Universidad Pedagógica Nacional?

Solución:

B: Profesor egresado de la Universidad Pedagógica

C: Profesor egresado de la Escuela Normal Superior

$B \cap C \neq \varnothing$ Dado que hay docentes que son egresados de una o de otra institución. Los sucesos no son mutuamente excluyentes.

Entonces:

$$P(B \text{ o } C) = P(B) + P(C) - P(B \cap C)$$

$$= \frac{12}{18} + \frac{6}{18} - \frac{5}{18} = \frac{13}{18} = 0.7222$$

$$= 72.22\%$$

El numerador de la expresión $\frac{5}{18}$ lo calculamos como en los casos anteriores, analizando en la tabla qué profesores cumplen la condición de ser egresados de la Escuela Normal Superior y de la Universidad Pedagógica Nacional. De acuerdo con la tabla, podemos ver que las personas que asistieron a una y a otra institución se ubican en los renglones M2, M6, H1, H4, H7, cuya suma total es de 5, lo que corresponde al valor obtenido.

Pregunta 4

¿Cuál es la probabilidad de que el profesor sea egresado de la Escuela Nacional de Maestros o de una escuela normal privada?

Solución:

A: Profesor egresado de la Nacional de Maestros

D: Profesor egresado de una escuela normal privada

$A \cap D = \varnothing$ En el personal docente que se está considerando no hay ningún docente que sea egresado de la Escuela Nacional de Maestros y que también sea de una escuela normal privada. Los sucesos son mutuamente excluyentes.

Entonces:

$$P(A \text{ o } D) = P(A) + P(D)$$

$$= \frac{12}{18} + \frac{6}{8} = \frac{18}{18} = 1$$

Que $P(A \text{ o } D) = 1$ señala que el suceso A o D es equivalente al suceso seguro. Recuerda que el suceso seguro es el que está representado por el conjunto universal U y $P(U) = 1$, y en este caso tenemos $P(A \text{ o } D) = 1$.

Este resultado significa que los sucesos A y D son complementarios porque $A \cap D = \varnothing$ y $A \cup D = U$. Es decir, un docente de la escuela que se está considerando es egresado de la Escuela Nacional de Maestros o bien, de una escuela normal privada.

Pregunta 5

¿Cuál es la probabilidad de que a un alumno le toque una profesora (mujer) o alguien egresado de la Escuela Normal Superior?

Solución:

C: Profesor egresado de la Normal Superior

E: Profesora (mujer)

$C \cap E \neq \varnothing$ Puesto que en el conjunto de docentes egresados de la Escuela Nacional Superior hay mujeres y en el conjunto de profesoras algunas son egresadas de la Escuela Normal Superior.

Entonces:

$$P(C \text{ o } E) = P(C) + P(E) - P(C \cap E)$$

$$= \frac{6}{18} + \frac{8}{18} - \frac{3}{18} = \frac{11}{18} = 0.6111$$

$$= 61.11\%$$

Nota: Explicación de cómo obtuvimos la expresión $\frac{8}{18}$, en especial el valor del numerador 8. Si observamos la tabla general podemos ver que son 8 mujeres las que intervienen en el problema, de M1 a M8.

Explicación de cómo obtuvimos la expresión $\frac{3}{18}$, en especial el valor del numerador 3. Analizando la tabla encontramos que las personas que corresponden a los renglones M2, M3 y M6 son profesoras egresadas de la Escuela Normal Superior.

Ley multiplicativa de la probabilidad

La probabilidad de que ocurran simultáneamente dos sucesos A y B se obtiene con el producto de sus probabilidades.

$$P(A \text{ y } B) = P(A) \cdot P(B)$$

Para aplicar la ley multiplicativa es necesario revisar si los sucesos involucrados son independientes o dependientes.

a) Sucesos independientes. Son aquellos en los que la ocurrencia de uno no efecta la probabilidad de que ocurra el otro.

Problema 17

Experimento aleatorio:

Se lanza un dado y se extrae una canica de una bolsa. En la bolsa hay 3 canicas: una roja, una azul y una verde. ¿Cuál es la probabilidad de que salga un número primo y una canica azul?

Solución:

Dado que cualquier resultado que aparezca en el dado no afecta la probabilidad del color de la canica que se extrae de la bolsa, ni viceversa, se dice que los sucesos son independientes.

A: {2, 3, 5} números primos

B: Se extrae una canica azul

$$P(A \text{ y } B) = P(A) \cdot P(B) = \left(\frac{3}{6}\right)\left(\frac{1}{3}\right) = \frac{3}{18} = 0.1666$$

$$= 16.66\%$$

Esto lo podemos comprobar contando de los resultados posibles los que son favorables al suceso A y B. Así:

$(A, 1)$ $(A, 2)$ $(A, 3)$ $(A, 4)$ $(A, 5)$ $(A, 6)$

$(R, 1)$ $(R, 2)$ $(R, 3)$ $(R, 4)$ $(R, 5)$ $(R, 6)$

$(V, 1)$ $(V, 2)$ $(V, 3)$ $(V, 4)$ $(V, 5)$ $(V, 6)$

Hemos marcado con una raya los resultados que satisfacen el problema, canica azul y número primo.

$$P(A \text{ y } B) = \frac{3 \text{ resultados favorables}}{18 \text{ resultados posibles}} = \frac{3}{18} = 0.1666$$

$$= 16.66\%$$

b) Sucesos dependientes. Son aquellos en los que la ocurrencia de uno afecta la probabilidad de que ocurra el otro.

En este caso, en la relación $P(A \text{ y } B) = P(A) \cdot P(B)$ se debe tener cuidado al calcular $P(B)$ porque esta probabilidad está afectada porque ya ocurrió el suceso A.

Problema 18

En una habitación hay 6 hombres y 3 mujeres. En la segunda habitación hay 5 hombres y 4 mujeres.

Calcula la probabilidad de elegir un hombre de la primera habitación y uno de la segunda.

Solución:

A: Hombre de la primera habitación

B: Hombre de la segunda habitación

Como son dos habitaciones, los eventos son independientes.

Entonces:

$$P(A \cap B) = P(A) \cdot P(B) =$$

$$\left(\frac{6}{9}\right)\left(\frac{5}{9}\right) = \frac{30}{81} = 37.03\%$$

Uso de las leyes aditivas y multiplicativas de la probabilidad

¿Cuál de ellas se debe aplicar en la solución de un problema?

Problema 19

En una escuela secundaria se realizará un sorteo para elegir a los representantes de cada grupo.

En el primer grado hay 50 alumnos, de ellos 25 son hombres.

En el segundo grado hay 60 alumnos, de ellos 20 son mujeres.

En el tercer grado hay 55 alumnos, de ellos 22 son hombres.

Consideremos los sucesos siguientes:

A: Todos los representantes son hombres

B: Todos los representantes son mujeres

C: Entre todos los representantes hay una mujer

Observa que ya no consideramos necesario citar el problema como un experimento aleatorio.

Solución:

Pregunta 1. Todos los representantes son hombres

Cada alumno tiene la misma probabilidad de salir seleccionado, pero lo que suceda en un grupo no cambia la probabilidad de lo que ocurra en el otro. Por ejemplo, si el representante del primer grado ya fue seleccionado, ésto no cambia el número de hombres ni el número de mujeres del segundo grado. Los sucesos son independientes.

Para calcular $P(A)$, se define:

A_1: Se selecciona a un hombre en el primer grado $P(A_1)=\frac{25}{50}=\frac{1}{2}$

A_2: Se selecciona a un hombre en el segundo grado $P(A_2)=\frac{40}{60}=\frac{2}{3}$

A_3: Se selecciona a un hombre en el tercer grado $P(A_3)=\frac{22}{55}=\frac{2}{5}$

Entonces:

$$P(A)=P(A_1 \text{ y } A_2 \text{ y } A_3)=P(A_1)\cdot P(A_2)\cdot P(A_3)$$
$$=\left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{2}{5}\right)=\frac{4}{30}=0.1333$$
$$=13.33\%$$

Pregunta 2. Todos los representantes son mujeres

Para calcular $P(B)$ se sigue un razonamiento análogo al citado.

Para calcular $P(B)$, se define:

B_1: Se selecciona a una mujer en el primer grado $P(B_1)=\frac{25}{50}=\frac{1}{2}$

B_2: Se selecciona a una mujer en el segundo grado $P(B_2)=\frac{20}{60}=\frac{1}{3}$

B_3: Se selecciona a una mujer en el tercer grado $P(B_3)\frac{33}{55}=\frac{3}{5}$

Entonces:

$$P(B) = P(B_1) \cdot P(B_2) \cdot P(B_3)$$

$$P(B) = \left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\left(\frac{3}{5}\right) = \frac{3}{30} = 0.1$$

$$= 10\%$$

Pregunta 3. Entre todos los representantes hay una mujer

Solución:

Las ternas (M, H, H) (H, M, H) (H, H, M) son resultados favorables al suceso que nos interesa. Los sucesos representados por cada terna son mutuamente excluyentes.

$$P(C) = P\,[(M, H, H) \text{ o } (H, M, H) \text{ o } (H, H, M)]$$

$$= P\,(M, H, H) + P(H, M, H) + P(H, H, M)$$

Entonces:

Calculamos cada probabilidad.

$$P(M, H, H) = P(M)P(H)P(H) = \left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = \frac{4}{30} = 0.1333$$

$$= 13.33\%$$

$$P(H, M, H) = P(H)P(M)P(H) = \left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\left(\frac{2}{5}\right) = \frac{2}{30} = 0.0666$$

$$= 6.66\%$$

$$P(H, H, M) = P(H)P(H)P(M) = \left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{3}{5}\right) = \frac{6}{30} = 0.20$$

$$= 20\%$$

Al calcular la probabilidad de cada terna, los sucesos que la componen son independientes, ya que, por ejemplo, si la representante del primer grado es mujer, esto no altera la probabilidad que un hombre salga seleccionado en el segundo grado; esto a su vez no cambia la probabilidad de la selección en el tercer grado.

Finalmente:

$$P(C) = P\,(M, H, H) + P(H, M, H) + P(H, H, M)$$

$$= 0.1333 + 0.0666 + 0.20 = 0.3999$$

$$= 39.99\%$$

Problema 20

En una caja hay 5 pelotas rojas, 8 blancas y 3 azules.

Consideremos los sucesos siguientes:

R: Se extrae pelota roja

B: Se extrae pelota blanca

C: Se extrae pelota azul

¿Cuál es la probabilidad de que al extraer una pelota sea roja o blanca?

¿Cuál es la probabilidad de que al extraer una pelota sea azul y blanca?

Pregunta 1. Probabilidad de extraer una pelota roja o blanca

Solución:

Los sucesos R y B son mutuamente excluyentes $R \cap B = \varnothing$

Entonces:

$$P(R \circ B) = P(R) + P(B) = \frac{5}{16} + \frac{8}{16} = \frac{13}{16} = 0.8125$$
$$= 81.25\%$$

Pregunta 2. Probabilidad de extraer una pelota azul y blanca

La probabilidad de extraer una pelota azul anula la probabilidad de que sea blanca; por ello, los sucesos son dependientes.

Solución:

A: Se extrae pelota azul

B: Se extrae pelota blanca

En este caso, el suceso que interesa es $A \cap B$, la intersección de los sucesos simples A, B.

$$P(A \cap B) = P(A) = \left(\frac{3}{16}\right)(0) = (0)$$

Del mismo modo, si la pelota es blanca se anula la probabilidad de que sea azul.

$$P(A \cap B) = P(A) \cdot P(B) = (0)\left(\frac{1}{2}\right) = 0$$

El suceso $A \cap B$ es un suceso imposible.

Observa lo siguiente: si en lugar de extraer una sola pelota como se plantea en el ejercicio, decimos que se extraen dos pelotas, una primero y otra después, sin regresar la primera, entonces se plantea un problema como el siguiente:

Pregunta 3. Probabilidad de que al extraer dos pelotas una sea blanca y la otra roja

Solución:

B: Se extrae pelota blanca

R: Se extrae pelota roja

Si la primera pelota es blanca, se afecta la probabilidad de que la segunda sea roja porque ahora esta última se selecciona de 15 pelotas disponibles

y no de 16. Por ello, los sucesos B y R son dependientes, uno efecta la probabilidad del otro.

Como el problema no plantea un orden para extraer el color y formar la pareja de pelotas, sólo dice que una vez obtenidas las dos pelotas se tenga una blanca y una roja. Así, se tiene que las parejas (B, R) y (R, B) son resultados favorables al suceso que nos interesa. En ambos casos extraer primero la pelota de un color afecta la probabilidad de sacar el otro, los sucesos R y B son dependientes al conformar una pareja de bolas de color.

Por otro lado, los sucesos (R, B) y (B, R) son mutuamente excluyentes, la probabilidad del suceso que plantea el problema es:

Solución:

$$P[(R, B) \text{ o } (B, R)] = P(R, B) + P(B, R) = P(R) \cdot P(B) + P(B) \cdot P(R)$$

$$= \left(\frac{5}{16}\right)\left(\frac{8}{15}\right) + \left(\frac{8}{16}\right)\left(\frac{5}{15}\right) = \frac{40}{240} + \frac{40}{240}$$

$$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = 0.3333$$

$$= 33.33\%$$

Pregunta 4. Probabilidad de que al realizar el suceso la primera pelota sea blanca y la segunda roja

Solución:

$$P(B \cap R) = P(B) \cdot P(R)$$

$$= \left(\frac{8}{16}\right)\left(\frac{5}{15}\right) = \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{1}{6} = 0.1666$$

$$= 16.66\%$$

Problema 21

La probabilidad de que un jugador de tenis gane un torneo es de $\frac{2}{5}$ y la probabilidad de que gane otro de los jugadores es de $\frac{1}{4}$. Calcula cuál es la probabilidad de que uno o el otro tenista gane el torneo.

Solución:

Dado que los sucesos son mutuamente excluyentes, la probabilidad es la suma de las probabilidades.

$$P(A) = \frac{2}{5}$$

$$P(B) = \frac{1}{4}$$

Con la relación:

$$P(A \cup B) = P(A) + P(B)$$

$$= \frac{2}{5} + \frac{1}{4} = \frac{8+5}{20}$$

$$= \frac{13}{20}$$

Problema 22

Tres equipos de fútbol A, B y C intervienen en un torneo; si A y B tienen la misma probabilidad de ganar y es tres veces la de C, calcula la probabilidad de que B o C ganen el torneo.

Solución:

$P(C) = x$

$P(A) = 3x$

$P(B) = 3x$

$x + 3x + 3x = 1$

$7x = 1$

$x = \frac{1}{7}$

$P(C) = \frac{1}{7}$

$P(A) = \frac{3}{7}$

$P(B) = \frac{3}{7}$

$(B \circ C) = (B \cup C)$

$P(B \cup C) = P(B) + P(C)$

$P(B \cup C) = \frac{3}{7} + \frac{1}{7}$

$= \frac{4}{7}$

Problema 23

La probabilidad de que un equipo de béisbol gane su primer juego es de $\frac{1}{4}$ y la probabilidad de que gane su segundo juego es de $\frac{1}{2}$. Calcula la probabilidad de que el equipo gane por lo menos uno de sus dos primeros juegos de un torneo si la probabilidad de que gane ambos es de $\frac{1}{5}$.

Solución:

A: El equipo gana el segundo juego

B: El equipo gana el primer juego

Como son sucesos no excluyentes entre sí, aplicamos la relación:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ con:

$P(A) = \frac{1}{2}$

$$P(B)=\frac{1}{4}$$

$$P(A\cap B)=\frac{1}{5}$$

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

$$=\frac{1}{2}+\frac{1}{4}-\frac{1}{5}=\frac{10+5-4}{20}$$

$$=\frac{11}{20}$$

Problema 24

A y B son sucesos en los que:

$$P(A)=\frac{3}{8}$$

$$P(B)=\frac{1}{2}$$

$$P(A\cap B)=\frac{1}{3}$$

Calcular:

a) $P(A\cup B)$

b) $P(A')$

c) $P(B')$

Solución:

a) Como A y B son sucesos no excluyentes entre sí, aplicamos la relación:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

$$=\frac{3}{8}+\frac{1}{2}-\frac{1}{3}=\frac{9+12-8}{24}$$

$$=\frac{13}{24}$$

b) $P(A')=1-P(A)$

$$=1-\frac{3}{8}=\frac{8}{8}-\frac{3}{8}$$

$$=\frac{5}{8}$$

La probabilidad de un suceso seguro es 1, el complemento de $P(A')=1-P(A)$

c) $P(B') = 1 - P(B)$

$$= 1 - \frac{1}{2} = \frac{2}{2} - \frac{1}{2}$$

$$= \frac{1}{2}$$

Probabilidad de una diferencia

La probabilidad de una diferencia se aplica cuando se quiere obtener la probabilidad de que un suceso determinado ocurra y que simultáneamente otro suceso, también determinado, no ocurra.

Se expresa así:

$$P(A - B) = P(A) - P(A \cap B)$$

Esta relación se conoce como *ley general de sustracción de probabilidades*. También se utilizan las relaciones:

$$P(A - B) = P(A) - P(A \cap B) = P(A' \cap B) = P(B \cap A) \quad \textbf{(1)}$$

$$P(A - B) = P(A) - P(A \cap B) \quad \textbf{(2)}$$

$$P(A') - P(A' \cap B) = P(A' \cap B') = P(A' - B) \quad \textbf{(3)}$$

Problema 25

En una urna hay pelotas rojas numeradas del 1 al 10 y pelotas azules numeradas del 1 al 5. ¿Cuál es la probabilidad de que al extraer al azar una pelota sea roja y no tenga el número 5?

Solución:

A: Se extrae pelota roja

B: Sale el número 5

El suceso que nos interesa es $A - B$. Aplicamos la relación:

$P(A - B) = P(A) - P(A \cap B)$

Con

$$P(A) = \frac{10}{15}$$

$$P(A \cap B) = \frac{1}{15}$$

Porque sólo hay una pelota roja marcada con el número 5 del total de las 15 pelotas.

Por lo tanto:

$$P(A - B) = \frac{10}{15} - \frac{1}{15} = \frac{9}{15}$$

$$= 0.6000$$

$$= 60\%$$

Otro procedimiento para resolver este problema es revisando la probabilidad como frecuencia relativa, así:

Solución:

Espacio muestral

$S = \{R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8, R_9, R_{10}, A_1, A_2, A_3, A_4, A_5\}$ son 15 sucesos

A: La pelota es roja y no es el número 5

$A = \{R_1, R_2, R_3, R_4, R_6, R_7, R_8, R_9, R_{10}\}$ son 9 (no está R_5)

$P(A) = \dfrac{\text{casos favorables}}{\text{casos posibles}} = \dfrac{9}{15} = 0.60 = 60\%$ es el mismo resultado.

El resultado que obtuvimos con $P(A \cap B) = P(A) \cdot P(B) = \dfrac{1}{15}$

También lo podemos obtener con las siguientes relaciones:

$P(A - B) = P(A \cap B') = P(B' \cap A)$

Problema 26

La probabilidad de que Antonio gane un juego de tenis es de $\dfrac{2}{5}$ y la probabilidad de que Juan gane es de $\dfrac{1}{4}$. ¿Cuál es la probabilidad de que Antonio gane el torneo en que participa si en el juego final se enfrenta a Juan?

Solución:

Sucesos

A: Gane Antonio

B: Gane Juan

El suceso que nos interesa es que Antonio gane y simultáneamente que Juan pierda. Por tanto, aplicamos la siguiente relación:

$P(A - B) = P(A) - P(A \cap B)$. Con $P(A) = \dfrac{2}{5}$, ahora es necesario calcular $P(A \cap B)$

Como A y B son sucesos dependientes aplicamos la siguiente relación:

$$P(A \cap B) = P(A \text{ y } B) = P(A) \cdot P(B)$$

$$= \frac{2}{5}\left(\frac{1}{4}\right) = \frac{2}{20}$$

$$P(A - B) = P(A) - P(A \cap B)$$

$$= \frac{2}{5} - \frac{2}{20} = \frac{8-2}{20} = \frac{6}{20} = 0.3$$

$$= 30\%$$

Otro procedimiento para resolver este problema consiste en aplicar la siguiente relación:

$P(A - B) = P(A) - P(A \cap B) = P(A \cap B')$

$P(A - B) = P(A \cap B')$

Con

$$P(A) = \frac{2}{5}$$

$$P(B') = 1 - \frac{1}{4} = \frac{4}{4} - \frac{1}{4} = \frac{3}{4}$$

Sustituimos en:

$$P(A - B) = P(A \cap B')$$
$$= P(A) \cdot P(B)$$
$$= \left(\frac{2}{5}\right)\left(\frac{3}{4}\right) = \frac{6}{20} = 0.3$$
$$= 30\%$$

Problema 27

A y B son sucesos donde:

$$P(A) = \frac{3}{4}$$

$$P(B) = \frac{3}{8}$$

$$P(A \cap B) = P(B \cap A) = \frac{1}{8}$$

Calcular:

a) $P(A' \cap B)$

b) $P(A \cap B')$

c) $P(A \cup B)$

d) $P(A' \cap B')$

Solución:

Recordando las relaciones de la probabilidad de una diferencia tenemos:

$P(A - B) = P(A) - P(A \cap B)$ **(1)**

$P(A - B) = P(A \cap B') - P(B' \cap A)$ **(2)**

a) De **(2)** intercambiamos A y B:

$P(B - A) = P(B \cap A') - P(A' \cap B)$ **(3)**

De **(1)** intercambiamos A y B:

$P(B - A) = P(B) - P(A \cap B)$ **(4)**

De **(3)** y **(4)**:

$$P(A' \cap B) = P(B) - P(A \cap B)$$

$$= \frac{3}{8} - \frac{1}{8} = \frac{2}{8} = \frac{1}{4} \qquad \textbf{(5)}$$

b) De **(2)** $P(A - B) = P(A \cap B')$

De **(1)** $P(A - B) = P(A) - P(A \cap B)$

Entonces:

$$P(A \cap B') = P(A) - P(A \cap B)$$

$$= \frac{3}{4} - \frac{1}{8} = \frac{5}{8} \qquad \textbf{(6)}$$

c) Sustituimos A por A'en **(2)**:

$$P(A' - B) = P(A' \cap B') \qquad \textbf{(7)}$$

Sustituimos A por A'en **(1)**:

$$P(A' - B) = P(A') - P(A' \cap B) \qquad \textbf{(8)}$$

De **(7)** y **(8)** obtenemos:

$$P(A' \cap B') = P(A') - P(A' \cap B) \qquad \textbf{(9)}$$

$$P(A') = 1 - P(A) = 1 - \frac{3}{4} = \frac{1}{4} \qquad \textbf{(10)}$$

De sustituir **(5)** y **(10)** en **(9)** obtenemos:

$$P(A' \cap B') = \frac{1}{4} - \frac{1}{4} = 0 \qquad \textbf{(11)}$$

d) $P(A' \cup B) = P(A') + P(B) - P(A' \cap B)$ **(12)**

Sustituyendo **(10)** y **(5)** en **(12)** tenemos:

$$P(A' \cup B) = \frac{1}{4} + \frac{3}{8} - \frac{1}{4} = \frac{3}{8}$$

Ventaja de un suceso

Se define como el cociente de la probabilidad de que el suceso ocurra entre la probabilidad de que no ocurra.

Aun cuando en el enunciado del problema los datos se citen en forma decimal o como tanto por ciento, las operaciones para resolver el problema se manejan como números racionales o como una proporción.

Problema 28

En una carrera de autos la probabilidad de que Pedro gane es de $\frac{1}{5}$. Calcula la ventaja que tiene sobre el resto de los competidores.

Solución:

$$\text{Ventaja} = \frac{\text{Probabilidad de ganar}}{\text{Probabilidad de perder}} = \frac{\frac{1}{5}}{\frac{4}{5}} = \frac{5}{20}$$

$$= \frac{1}{4}$$

Se dice entonces que Pedro tiene una ventaja de ganar de 1 a 4.

De otra manera, si lo que sabemos es que Pedro tiene una ventaja de 1 a 4, podemos averiguar cuál es la probabilidad de que Pedro gane.

Solución:

Sucesos

A: Pedro gane

B: Pedro no gane

$$\text{Ventaja} = \frac{P(A)}{P(A')}$$

$$\frac{1}{4} = \frac{P(A)}{P(A')}$$

Despejamos:

$$P(A) = \frac{1}{4} \cdot P(A') \text{ como } P(A') = 1 - P(A)$$

Sustituimos:

$$P(A) = \frac{1}{4}\left[1 - P(A)\right]$$

Operaciones:

$$P(A) = \frac{1}{4} - \frac{1}{4}P(A) \text{ el coeficiente de } P(A) \text{ es 1, que es igual a } \frac{4}{4}$$

Despejamos:

$$\frac{4}{4}P(A) + \frac{1}{4}P(A) = \frac{1}{4}$$

$$\frac{5}{4}P(A) = \frac{1}{4}$$

$$P(A) = \frac{\frac{1}{4}}{\frac{5}{4}} = \frac{4}{20}$$

$$= \frac{1}{5}$$

Nota: Un procedimiento sencillo de obtener la *probabilidad* de un suceso A si se sabe que A tiene una ventaja de x a y es la siguiente:

Ventaja: x a y, que también se puede escribir como x: y que es lo mismo que $\frac{x}{y}$.

La probabilidad de que Pedro gane es de $\frac{x}{x+y}$

Sustituimos con $x=1, y=4$

$$P(A)=\frac{1}{1+4}$$

$\frac{1}{5}$, que es el mismo resultado anterior.

Problema 29

En un campeonato de fútbol, Mario tiene dos equipos favoritos A y B; la ventaja del equipo A es de 2 a 3, mientras que la ventaja del equipo B es de 2 a 5. ¿Cuál es la probabilidad de que se corone como campeón alguno de los dos equipos favoritos de Mario?

Solución:

$$V(A): 2 \text{ a } 3 = 2:3 = \frac{2}{3}$$

$$P(A)=\frac{2}{2+3}=\frac{2}{5}$$

$$V(B): 2 \text{ a } 5 = 2:5 = \frac{2}{5}$$

$$P(B)=\frac{2}{2+5}=\frac{2}{7}$$

Con los valores obtenidos de $P(A)$ y $P(B)$ calculamos $P(A \text{ o } B) = P(A \cup B)$. Como los sucesos A y B son mutuamente excluyentes, aplicamos la relación $P(A \cup B) = P(A) + P(B)$

$$=\frac{2}{5}+\frac{2}{7}=\frac{14+10}{35}=\frac{24}{35}$$

$$=0.6857$$

$$=68.57\%$$

Resumen

Probabilidad como frecuencia relativa

$$\text{Probabilidad } S = \frac{\text{Número de veces que el suceso } E \text{ ocurre}}{\text{Total de sucesos realizados}} = \frac{h}{n}$$

Probabilidad con base en sucesos compuestos. Probabilidad axiomática

En los sucesos compuestos se pueden aplicar tres leyes en función del resultado que se esté calculando.

Ley aditiva

Se utiliza cuando se quiere obtener la probabilidad de que ocurra el suceso A o el suceso B.

Para aplicar esta ley es necesario saber si los sucesos A y B son mutuamente excluyentes o no.

a) Si A y B son *mutuamente excluyentes* se aplica la relación $P(A \text{ o } B) = P(A) + P(B)$. Recuerda que en $(P \text{ o } B) = (A \cup B)$ deberás usar la notación que indique tu profesor.

b) Si A y B no son *mutuamente excluyentes* se aplica la relación $P(A \text{ o } B) = P(A) + P(B) - P(A \cap B)$.

Ley multiplicativa

Se utiliza cuando se necesita saber cuál es la probabilidad de que dos sucesos A y B ocurran simultáneamente.

Para aplicar esta ley es necesario saber si los sucesos A y B son independientes o dependientes.

a) Si A y B son *independientes* se aplica la relación $P(A \text{ y } B) = P(A) \cdot P(B)$.

Recuerda que en $P(A \text{ y } B) = (A \cap B)$ usarás la notación que indique tu profesor.

b) Si A y B son dependientes se aplica la misma relación anterior, sólo que para calcular la $P(B)$ debe hacerse tomando en cuenta que ya sucedió A.

La relación $P(A \text{ y } B) = P(A) \cdot P(B)$ para A y B dependientes, su aplicación es un caso particular de la probabilidad condicional que se estudiará más ampliamente en el capítulo siguiente.

Ley de la diferencia

Se utiliza cuando es necesario que ocurra el suceso A y simultáneamente no ocurra el suceso B. Se aplica la siguiente relación.

$$P(A - B) = P(A) - P(A \cap B) \qquad \textbf{(1)}$$

También:

$$P(A - B) = P(A) - P(A \cap B) = P(A' \cap B) = P(B \cap A') \qquad \textbf{(2)}$$

$$P(A') - P(A' \cap B) = P(A' \cap B') = P(A' - B) \qquad \textbf{(3)}$$

Con la relación **(2)** obtuvimos:

$$P(B - A) = P(B) - P(B \cap A) = P(B' \cap A) = P(A \cap B') \qquad \textbf{(a)}$$

$$P(A' - B) = P(A') - P(A' \cap B) = P[(A')' \cap B] = P(A \cap B)$$

$$= P(B \cap A) \qquad \textbf{(b)}$$

$$P(A - B') = P(A) - P(A \cap B') = P(A' \cap B') = P(B' \cap A') \qquad \textbf{(c)}$$

Además:

$$P(A \cup B') = P(A) + P(B') - P(A \cap B') \quad \textbf{(d)}$$

$$P(A' \cup B) = P(A') + P(B) - P(A' \cap B) \quad \textbf{(e)}$$

Probabilidad condicional

Consideraciones generales

La probabilidad condicional se aplica en el cálculo de un suceso cuando se sabe que ha ocurrido otro con el cual se relaciona, es decir, los sucesos son dependientes.

Por ejemplo: Calcula cuál es la probabilidad de que baje la temperatura en el Valle de México si hay un frente frío en el Golfo de México.

Sean A y B dos sucesos dependientes, tales que $P(A) > 0$.

Para expresar la probabilidad de B dado que A ha ocurrido, se expresa $P(B \mid A)$.

Del mismo modo, si $P(B) > 0$.

Para señalar la probabilidad de A dado que B ha ocurrido, se expresa $P(A \mid B)$.

Consideremos $P(B \mid A)$.

La probabilidad de $(B \mid A)$ se realiza en un nuevo espacio muestral, que es un subconjunto del espacio muestral original S. Es decir, el espacio muestral original S se ve modificado porque ya ocurrió el suceso A.

Para ilustrar la razón por la que en la probabilidad condicional el espacio muestral orginal se modifica, analicemos el problema siguiente:

Problema 30

En una urna hay 6 pelotas azules numeradas del 1 al 6. En otra urna hay seis pelotas rojas también numeradas del 1 al 6. ¿Cuál es la probabilidad de que al sacar dos pelotas, una de cada urna, la suma de los números sea mayor que seis si ya sabemos que una pelota azul salió con un número divisible entre dos?

Solución:

El espacio muestral original S es el siguiente:

$$\begin{array}{cccccc} A_1R_1 & A_1R_2 & A_1R_3 & A_1R_4 & A_1R_5 & A_1R_6 \\ A_2R_1 & A_2R_2 & A_2R_3 & A_2R_4 & A_2R_5 & A_2R_6 \\ A_3R_1 & A_3R_2 & A_3R_3 & A_3R_4 & A_3R_5 & A_3R_6 \\ A_4R_1 & A_4R_2 & A_4R_3 & A_4R_4 & A_4R_5 & A_4R_6 \\ A_5R_1 & A_5R_2 & A_5R_3 & A_5R_4 & A_5R_5 & A_5R_6 \\ A_6R_1 & A_6R_2 & A_6R_3 & A_6R_4 & A_6R_5 & A_6R_6 \end{array}$$

El nuevo espacio muestral es aquel en el que la pelota azul salió con un número divisible entre dos, el cual se conforma de las parejas del segundo, cuarto y sexto renglón. En este nuevo espacio muestral se calcula la probabilidad de que la suma de los números sea mayor que seis.

Las parejas marcadas en este espacio son los sucesos favorables, mismo que se indica a continuación:

A_2R_1	A_2R_2	A_2R_3	A_2R_4	A_2R_5	A_2R_6
A_4R_1	A_4R_2	A_4R_3	A_4R_4	A_4R_5	A_4R_6
A_6R_1	A_6R_2	A_6R_3	A_6R_4	A_6R_5	A_6R_6

Si tenemos:

A: En la pelota azul se obtiene un número divisible entre dos.

B: La suma de los números es mayor que seis.

Entonces, la probabilidad de que suceda B, dado que ya ocurrió A es:

$$P(B \mid A) = \frac{12}{18} = 0.6666$$
$$= 66.66\%$$

El numerador de la expresión $\frac{12}{18}$ lo obtuvimos sumando los sucesos favorables en el espacio circulado y el denominador es la suma de todos los sucesos del nuevo espacio muestral.

Como no siempre es posible escribir todos los sucesos simples de un espacio muestral, para obtener el resultado se aplica la relación siguiente:

Si A y B son dos sucesos dependientes, tales que $P(A) > 0$, la probabilidad de que ocurra B cuando ya sucedió A es:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Despejando:

$$P(A \cap B) = P(A)\ P(B \mid A)$$

A continuación aplicamos esta relación al problema anterior.

A: En la pelota azul sale un número divisible entre dos.

B: La suma de los números es mayor que seis.

El suceso $A \cap B$ y el suceso A suceden en el espacio muestral original S.

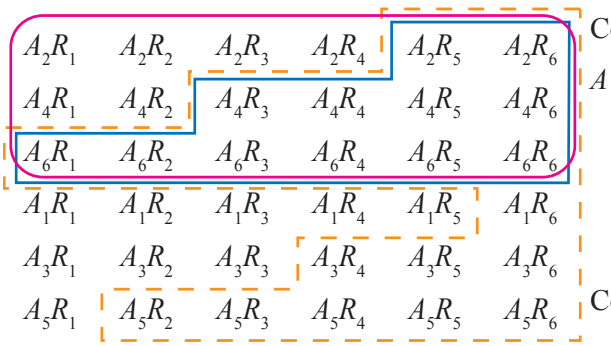
$A \cap B$: En la pelota azul se obtiene un número divisible entre dos y la suma de los números es mayor que seis.

Se señala a continuación:

Las parejas del conjunto A (resaltados con una línea rosa).

Las parejas del conjunto B (resaltados con una línea punteada)

Las parejas del espacio muestral S que están en la intersección de A con B (resaltadas con una línea azul).



Número de sucesos favorables para $A = 18$ (resaltado con una línea rosa).

Número de sucesos favorables para $B = 21$ (resaltados con una línea punteada).

Número de sucesos favorables para $A \cap B = 12$ (resaltado con una línea azul).

Entonces:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{12}{36}}{\frac{18}{36}} = \frac{12}{18} = 0.6666$$

$$= 66.66\%$$

Con la relación de la probabilidad condicional obtuvimos el mismo resultado que cuando se trabajó con la redefinición del espacio muestral original, lo citamos como nuevo espacio muestral.

Propiedades

La probabilidad condicional satisface las propiedades de la frecuencia relativa en la forma siguiente:

- **1.** Para los sucesos $A \cap B$.

  $0 \leq P(A \cap B) \leq 1$

  Es decir, la probabilidad de $A \cap B$ es mayor o igual que cero, y menor que uno.

- **2.** $P(A \cap B) = 1$ si y sólo si $A \cap B$ ocurre en las n repeticiones.

- **3.** $P(A \cap B) = 0$ si y sólo si $A \cap B$ nunca ocurre en las repeticiones.

Además cumple con las siguientes:

a) Para sucesos dependientes:

Para tres sucesos dependientes A_1 , A_2 , A_3 se tiene:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2)$$

b) Para sucesos independientes, tales que $P(A) > 0$ y $P(B) > 0$ se tiene que:

$P(A \cap B) = P(A) \cdot P(B)$

c) Además, si A y B son sucesos independientes, se tiene la siguiente relación:

$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$

A continuación, nos referimos a cada una de ellas.

Para sucesos dependientes

Sean tres sucesos A_1 , A_2 , A_3, señalamos que se tiene:

$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2)$

La probabilidad de que A_1 , A_2 y A_3 ocurran es el producto de la probabilidad de que ocurra A_1, por la probabilidad de que ocurra A_2 si ya sucedió A_1, por la probabilidad de que ocurra A_3 sabiendo que A_1 y A_2 han ocurrido.

Problema 31

Si la probabilidad de que ocurra el suceso A es $\frac{1}{2}$ igual a la probabilidad de que suceda B, y se sabe que la probabilidad de que ocurra B una vez que ya sucedió A es de $\frac{2}{3}$. Calcula $P(A \text{ y } B)$ y $P(A \text{ o } B)$.

Como A y B son sucesos dependientes aplicamos la siguiente relación:

$P(A \cap B) = P(A) \cdot P(B)$

$P(A) = \frac{1}{2}$

$P(B) = \frac{2}{3}$

$P(A \cap B) = P(A) \cdot P(B)$

$= \frac{1}{2}\left(\frac{2}{3}\right) = \frac{2}{6}$

$= \frac{1}{3} = 0.3333$

$= 33.33\%$

Como A y B no son mutuamente excluyentes, aplicamos la siguiente relación:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(A)=\frac{1}{2}$$

$$P(B)=\frac{1}{2}$$

$$P(A\cap B)=\frac{1}{3}$$

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

$$=\frac{1}{2}+\frac{1}{2}-\frac{1}{3}=\frac{3+3-2}{6}=\frac{4}{6}$$

$$=\frac{2}{3}=0.6666$$

$$=66.66\%$$

Problema 32

En un grupo se realizará un sorteo para elegir a los tres alumnos que se harán cargo de la ceremonia escolar. En el grupo hay 24 hombres y 12 mujeres. ¿Cuál es la probabilidad de que el grupo de representantes esté conformado de las maneras siguientes?

A. Sean tres hombres

B. Sean dos hombres y una mujer

C. Sean dos mujeres y un hombre

D. Sean tres mujeres

Solución:

A. Sean tres hombres.

Obtenemos el resultado con la relación de la probabilidad condicional.

$A=A_1\cap A_2\cap A_3$

A_1: el primer alumno seleccionado es hombre

A_2: el segundo alumno seleccionado es hombre

A_3: el tercer alumno seleccionado es hombre

$P(A)=(A_1\cap A_2\cap A_3)=P(A_1)\cdot P(A_2\mid A_1)\cdot P(A_3\mid A_1\cap A_2)$

Calculamos por separado:

$$P(A_1)=\frac{24}{36} \qquad (1)$$

$$P(A_2\mid A_1)=\frac{P(A_1\cap A_2)}{P(A_1)} \qquad (2)$$

$$P(A_2\mid A_1)=\frac{23}{35} \qquad (3)$$

Nota: $P(A_2 \mid A_1) = \frac{23}{35}$ porque sabemos que ya ocurrió A_1 (el primer alumno seleccionado es hombre), entonces sólo quedan 23 hombres de un nuevo espacio muestral de 35 alumnos.

Sustituyendo **(1)** y **(3)** en **(2)**:

$$\frac{23}{35} = \frac{P(A_1 \cap A_2)}{\frac{24}{36}}$$

$$P(A_1 \cap A_2) = \left(\frac{23}{35}\right)\left(\frac{24}{36}\right) = \frac{552}{1\,260} = \frac{46}{105} \qquad \textbf{(4)}$$

$$P(A_3 \mid A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \qquad \textbf{(5)}$$

$$P(A_3 \mid A_1 \cap A_2) = \frac{22}{34} \qquad \textbf{(6)}$$

Sabemos que ya ocurrió $A_1 \cap A_2$, es decir el primero fue hombre y el segundo también fue hombre. Ahora sólo quedan 22 hombres en un nuevo espacio muestral de 34 personas.

Sustituyendo **(4)** y **(6)** en **(5)**

$$\frac{22}{34} = \frac{P(A_1 \cap A_2 \cap A_3)}{\frac{46}{105}}$$

$$P(A_1 \cap A_2 \cap A_3) = \frac{22}{34}\left(\frac{46}{105}\right) = \frac{1\,012}{3\,570} = .2834$$

$$= 28.34\%$$

B. Sean dos hombres y una mujer.

Existen tres ternas que son favorables al evento:

(M, H, H), (H, M, H), (H, H, M)

$P(B) = P(M, H, H) + P(H, M, H) + P(H, H, M)$

Calculando por separado tenemos:

- $P(M \cap H \cap H) = P(M) \cdot P(H \mid M) \cdot P(H \mid M \cap H)$

$$= \frac{12}{36}\left(\frac{24}{35}\right)\left(\frac{23}{34}\right) = \frac{6\,624}{42\,840} = .1546$$

$$= 15.46\%$$

- $P(H \cap M \cap H) = P(H) \cdot P(M \mid H) \cdot P(H \mid H \cap M)$

$$= \frac{24}{36}\left(\frac{12}{35}\right)\left(\frac{23}{34}\right) = \frac{6\,624}{42\,840} = 15.46\%$$

- $P(H \cap H \cap M) = P(H) \cdot P(H \mid H) \cdot P(M \mid H \cap H)$

$$= \frac{24}{36}\left(\frac{23}{36}\right)\left(\frac{12}{35}\right) = \frac{6\,624}{42\,840} = 15.46\%$$

$$P(B) = 15.46\% + 15.46\% + 15.46\% = 46.38\%$$

Nota: $P(H \mid M) = \frac{24}{35}$. El 24 es porque hay 24 hombres, pero sólo hay 35 personas, ya que una de ellas ya fue elegida, una mujer.

C. Sean dos mujeres y un hombre.

Como en el caso anterior, existen tres casos favorables:

(M, M, H), (M, H, M), (H, M, M)

$P(C) = P(M \cap M \cap H) + P(M \cap H \cap M) + P(H \cap M \cap M)$

- $P(M \cap M \cap H) = P(M) \cdot P(M \mid M) \cdot P(H \mid M \cap M)$

$$= \frac{12}{36}\left(\frac{11}{35}\right)\left(\frac{24}{34}\right) = \frac{3\,168}{42\,840} = 7.39\%$$

- $P(M \cap H \cap M) = P(M) \cdot P(H \mid M) \cdot P(M \mid M \cap H)$

$$= \frac{12}{36}\left(\frac{24}{35}\right)\left(\frac{11}{34}\right) = \frac{3\,168}{42\,840} = 7.39\%$$

- $P(H \cap M \cap M) = P(H) \cdot P(M \mid H) \cdot P(M \mid H \cap M)$

$$= \frac{24}{36}\left(\frac{12}{35}\right)\left(\frac{11}{34}\right) = \frac{3\,168}{42\,840} = 7.39\%$$

$P(C) = 7.39\% + 7.39\% + 7.39\% = 22.17\%$

D. Sean tres mujeres.

- $P(M \cap M \cap M) = P(M) \cdot P(M \mid M) \cdot P(M \mid M \cap M)$

$$= \left(\frac{12}{36}\right)\left(\frac{11}{35}\right)\left(\frac{10}{34}\right)$$

$$= \frac{1\,320}{42\,840} = 3.08\%$$

Nota: Algunos autores expresan $P(A$ y $B)$ como $P(A \cap B)$. En este caso, el uso del símbolo $\cap$ (intersección) no debe interpretarse como en la teoría de conjuntos porque de esta intersección puede resultar el conjunto vacío y lo que esperamos es que se realicen dos sucesos simples:

A: se elige un hombre

B: se elige una mujer

A y B representan el deseo de que la pareja seleccionada esté formada por un hombre y por una mujer, y al usar la notación $A \cap B$ para representar este suceso, no debe confundirse con el hecho de que al seleccionar a una persona queremos que ésta sea al mismo tiempo hombre y mujer, lo cual no es posible.

Podemos generalizar el razonamiento para n sucesos dependientes A_1, A_2, A_3,... A_n, así:

$P(A_1 \cap A_2 \cap A_3,... \cap A_n)$

$P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_2 \cap A_1) ... \cdot P(A_n \mid A_1 \cap ... \cap A_{n-1})$

Esta relación la identificamos como la propiedad multiplicativa de la probabilidad.

Para sucesos independientes

Para dos sucesos A y B independientes, tales que $P(A) > 0$ y $P(B) > 0$ se tiene $P(A \cap B) = P(A) \cdot P(B)$, relación que obtuvimos del siguiente modo:

Como:

$P(B \mid A) = P(B)$, ya que la ocurrencia del suceso A no afecta la probabilidad de que suceda B.

Del mismo modo, $P(A \mid B) = P(A)$ y por ser sucesos independientes tenemos:

La relación de la probabilidad condicional

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

Despejamos:

$$P(A \cap B) = P(A) \cdot P(B \mid A)$$

Como $P(B \mid A) = P(B)$ obtenemos $P(A \cap B) = P(A) \cdot P(B)$

Ésta es la ley multiplicativa que analizamos en el capítulo dos.

Problema 33

En una urna hay 6 pelotas azules numeradas del 1 al 6 y en otra urna hay 6 pelotas rojas también numeradas del 1 al 6.

Experimento:

Extraer al azar una pelota de cada urna.

Sucesos:

A_3: En la pelota azul sale el número 3

R_2: En la pelota roja sale el número 2

¿Cuál es la probabilidad de que ocurran A_3 y R_2?

Observa que el experimento consiste en extraer una pelota azul y una roja; y no lo que uno podría pensar equivocadamente que $A_3 \cap R_2 = \varnothing$ si se piensa que sacar pelota azul no permite que salga pelota roja, idea fuera del contexto del problema.

Lo que se tiene es que A_3 y R_2 son sucesos independientes, por lo que la probabilidad de que ocurra uno no cambia la probabilidad de la ocurrencia del otro.

Solución:

$$P(A_3 \cap R_2) = P(A_3) \cdot P(R_2) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}$$

La probabilidad de que ocurran los sucesos A_3 y R_2 es de $\frac{1}{36}$.

Si queremos resolver el problema con la expresión de la probabilidad condicional para calcular $P(A_3 \cap R_2)$ tenemos:

Solución:

$$P(R_2 \mid A_3) = \frac{P(A_3 \cap R_2)}{P(A_3)}$$

Despejando:

$$P(A_3 \cap R_2) = P(R_2 \mid A_3) \cdot P(A_3)$$

Calculamos:

$P(R_2 \mid A_3) = \frac{1}{6}$ en el espacio muestral definido por A_3

$P(A_3) = \frac{6}{36}$ en el espacio muestral original

$$P(A_3 \cap R_2) = \left(\frac{1}{6}\right)\left(\frac{6}{36}\right) = \frac{1}{36}$$

La probabilidad de que ocurran los sucesos A_3 y R_2 es de $\frac{1}{36}$.

La relación $P(A \cap B) = P(A) \cdot P(B)$ se puede generalizar a n sucesos A_1, A_2, A_3,…, A_n independientes. Así:

$$P(A_1 \cap A_2 \cap A_3 \ldots A_n) = P(A_1) \cdot P(A_2) \ldots P(A_n)$$

Si A y B son sucesos independientes, cumplen la siguiente relación:

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$$

Problema 34

Una maestra tiene cuatro pelotas de colores diferentes y tres libros de temas diferentes. Ella quiere regalar a sus mejores alumnos un libro y una pelota. Para repartirlos hace un sorteo.

José quiere la pelota amarilla y el libro de aviones, pero se conformaría con que al menos uno de los dos regalos que quiere le toque a él. ¿Qué probabilidad tiene de que se cumpla su deseo?

Solución:

A. Obtiene pelota amarilla

B. Obtiene el libro de aviones

El suceso que le conviene a José es $A \cup B$

Como A y B son sucesos independientes, aplicamos:

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$$

$$P(A \cup B) = \left(\frac{1}{4} + \frac{1}{3}\right) - \left(\frac{1}{4}\right)\left(\frac{1}{3}\right) = \frac{7}{12} - \frac{1}{12}$$

$$= \frac{1}{2}$$

El resultado anterior también lo podemos obtener con la relación como frecuencia relativa.

Solución:

Para establecer el espacio muestral numeramos las pelotas del 1 al 4 y el 1 le corresponde a la pelota amarilla; de la misma manera, numeramos los libros del 1 al 3 y el 1 le corresponde al libro de aviones.

Solución:

El espacio muestral es:

P_1L_1	P_1L_2	P_1L_3
P_2L_1	P_2L_2	P_2L_3
P_3L_1	P_3L_2	P_3L_3
P_4L_1	P_4L_2	P_4L_3

Del espacio muestral se tiene que hay 6 sucesos favorables de los 12 posibles, por lo tanto:

$$P(A \cup B) = \frac{6}{12} = \frac{1}{2}$$

El resultado también se puede obtener con la ley aditiva de la probabilidad, la cual establece que para el caso de que A y B no sean mutuamente excluyentes $(A \cup B) = \varnothing$, entonces:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= \frac{1}{4} + \frac{1}{3} - \frac{1}{12} = \frac{3+4-1}{12} = \frac{6}{12}$$

$$= \frac{1}{2}$$

Problemas resueltos

Problema 35

De un lote de 15 camisas, 4 están defectuosas. Si se toman al azar 3 artículos del lote, uno tras otro, calcula la probabilidad de que los tres se encuentren en buen estado.

Solución:

A_1: la primera camisa está en buen estado

A_2: la segunda camisa está en buen estado

A_3: la tercera camisa está en buen estado

El suceso que nos interesa es $A_1 \cap A_2 \cap A_3$; los sucesos A_1, A_2 y A_3 son dependientes porque al extraer la primera camisa en buen estado, cambia el número de casos favorables para extraer la segunda camisa y también cambia el total de camisas disponibles, entonces:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2)$$

$$= \left(\frac{11}{15}\right)\left(\frac{10}{14}\right)\left(\frac{9}{13}\right) = \frac{990}{2\,730} = 0.36$$

$$= 36\%$$

Problema 36

En una escuela de enseñanza media superior, 20% de los alumnos reprobó matemáticas, 25% física y 5% ambas materias. Si se selecciona un alumno al azar:

a) Si reprobó física, ¿cuál es la probabilidad de que haya reprobado matemáticas?

b) Si reprobó matemáticas, ¿cuál es la probabilidad de que haya reprobado física?

c) ¿Cuál es la probabilidad de que haya reprobado física o matemáticas?

Solución:

$$P(M) = 20\% = \frac{1}{5}$$

$$P(F) = 25\% = \frac{1}{4}$$

$$P(M \cap F) = 5\% = \frac{1}{20}$$

Las cantidades citadas en tanto por ciento se pasaron a números racionales

a) $P(F \mid M) = \dfrac{P(F \cap M)}{P(M)} = \dfrac{\frac{1}{20}}{\frac{1}{5}} = \dfrac{5}{20} = \dfrac{1}{4}$

b) $P(M \mid F) = \dfrac{P(M \cap F)}{P(F)} = \dfrac{\frac{1}{20}}{\frac{1}{4}} = \dfrac{4}{20} = \dfrac{1}{5}$

c) $P(M \cup F) = P(M) + P(F) - P(M \cap F)$

$$= \frac{1}{5} + \frac{1}{4} - \frac{1}{20} = \frac{4+5-1}{20} = \frac{8}{20} = \frac{2}{5}$$

Problema 37

Sean A y B dos sucesos donde:

$$P(A) = \frac{1}{2}$$

$$P(B) = \frac{1}{5}$$

$$P(A \cap B) = \frac{1}{3}$$

Determina:

a) $P(B \mid A)$

b) $P(A \cup B)$

c) $P(A' \mid B')$

Solución:

a) $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{\frac{1}{3}}{\frac{1}{2}} = \dfrac{2}{3}$

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$= \frac{1}{2} + \frac{1}{5} - \frac{1}{3} = \frac{15+6-10}{30}$$

$$= \frac{11}{30}$$

c) De la teoría de conjuntos (Leyes de Morgan) tenemos:

$A' \cap B' = (A \cup B)'$

Entonces:

$$P(A' \mid B') = \frac{P(A' \cap B')}{P(B')} = \frac{P(A \cup B)'}{1 - P(B)} = \frac{1 - P(A \cup B)}{1 - P(B)}$$

$$= \frac{1 - \frac{11}{30}}{\frac{5}{5} - \frac{1}{5}} = \frac{\frac{19}{30}}{\frac{4}{5}} = \frac{95}{120}$$

$$= \frac{19}{24}$$

Problema 38

Se lanza un dardo dos veces sobre un círculo que gira; el círculo está dividido en 6 sectores iguales numerados del 1 al 6. Calcula la probabilidad de obtener un 3 o un 4 en el primer lanzamiento y un 1, un 2, un 3 o un 5 en el segundo lanzamiento.

Solución:

A: sale un 3 o un 4 en el primer lanzamiento

B: sale un 1, un 2, un 3 o un 5 en el segundo lanzamiento

A y B son sucesos independientes, lo que ocurre en el primer lanzamiento no afecta la probabilidad de lo que ocurra en el segundo lanzamiento, entonces:

$$P(A \cap B) = P(A) \cdot P(B)$$

$$= \left(\frac{2}{6}\right)\left(\frac{4}{6}\right) = \frac{8}{36}$$

$$= \frac{2}{9}$$

Este tipo de problemas también lo podemos resolver en la forma siguiente:

Solución:

En el primer lanzamiento hay 6 maneras en que el dardo puede caer, lo que puede asociarse con cada una de las 6 maneras en que cae en el segundo lanzamiento, todas ellas igualmente posibles. Hay 6(6) = 36 maneras diferentes.

Cada una de las dos maneras en que A ocurre, las asociamos con cada una de las cuatro maneras en que B ocurre, así se obtiene 2(4) = 8 maneras en que tanto A como B ocurren.

Por lo tanto:

$$P(A \cap B) = \frac{8}{36} = \frac{2}{9}$$

Problema 39

En una escuela de enseñanza media superior, 40% de la población de alumnos mide más de 1.50 m, 25% pesa más de 52 kg y 15% mide más de 1.50 m y pesa más de 52 kg. Si se escoge al azar un alumno:

a) Si mide más de 1.50 m, calcula la probabilidad de que también pese más de 52 kg.

b) ¿Cuál es la probabilidad de que no mida más de 1.50 m ni pese más de 52 kg?

Solución:

A: mide más de 1.50 m

K: pesa más de 52 kg

$A \cap K$: mide más de 1.50 m y pesa más de 52 kg.

$$P(A) = 40\% = \frac{40}{100} = \frac{2}{5}$$

$$P(K) = 25\% = \frac{25}{100} = \frac{1}{4}$$

$$P(A \cap K) = 15\% = \frac{15}{100} = \frac{3}{20}$$

a) $$P(A \mid K) = \frac{P(A \cap K)}{P(K)} = \frac{\frac{3}{20}}{\frac{1}{4}} = \frac{12}{20}$$

$$= \frac{3}{5}$$

b) De la teoría de conjuntos tenemos:

$$(A \cup B)' = A' \cap B'$$

Entonces:

$$P(A' \cap K') = P(A \cup K)' = 1 - P(A \cup K)$$

$$= 1 - \frac{1}{2}$$

$$= \frac{1}{2}$$

Nota: $P(A \cup K) = P(A) + P(K) - P(A \cap K)$

$$= \frac{2}{5} + \frac{1}{4} - \frac{3}{20}$$

$$= \frac{10}{20} = \frac{1}{2}$$

Problema 40

Se extraen al azar 2 cartas de una baraja de 36. Determina la probabilidad de que ambas sean sotas si la carta se reemplaza y si no se reemplaza.

Solución:

a) La carta se reemplaza

$P(A)$ = sota en la primera extracción

$P(B)$ = sota en la segunda extracción

$$P(A \cap B) = P(A) \cdot P(B \mid A)$$

$$= \frac{4}{36}\left(\frac{4}{36}\right) = \frac{16}{1\,296}$$

$$= \frac{1}{81}$$

En la primera extracción hay 4 sotas en las 36 cartas, por eso $P(A) = \frac{4}{36}$; como la carta se reemplaza para la segunda extracción, vuelve a haber 4 sotas; por lo tanto, son eventos independientes:

$$P(B \mid A) = \frac{4}{36}$$

b) La carta no se reemplaza

$$P(A \cap B) = P(A) \cdot P(B \mid A)$$

$$= \frac{4}{36}\left(\frac{3}{35}\right) = \frac{12}{1\,260}$$

$$= \frac{1}{105}$$

En la primera extracción hay 4 sotas en las 36 cartas, por eso $P(A) = \frac{4}{36}$; para la segunda extracción únicamente quedan 3 sotas en 35 cartas, entonces:

$$P(B \mid A) = \frac{3}{35}$$

Este problema también lo podemos resolver en la forma siguiente:

a) La carta se reemplaza

La primera extracción se puede hacer en una de las 36 maneras posibles y de inmediato se regresa a la baraja. Entonces, la segunda sota también se puede extraer en una de las 36 maneras posibles, así que las dos cartas pueden extraerse de 36 (36) maneras, todas igualmente factibles.

En este caso, hay 4 maneras de obtener una sota en la primera extracción y 4 maneras de obtener una sota en la segunda, de donde el número de maneras de obtener una sota en la primera y segunda extracción es de 4 (4). Por lo tanto:

$$\text{Probabilidad (obtener 2 sotas)} = \frac{4(4)}{36(36)} = \frac{16}{1\,296} = \frac{1}{81}$$

b) La carta no se reemplaza

La primera carta puede extraerse de una de las 36 maneras posibles y como no se devuelve al mazo de cartas, la segunda carta puede extraerse de 35 maneras posibles. Por lo tanto, las dos cartas pueden extraerse en 36 (35) maneras, todas igualmente factibles.

Cuando la carta no se reemplaza, en la primera extracción hay 4 maneras de obtener una sota y 3 maneras de obtener una sota en la segunda. Así tenemos que la probabilidad para obtener una sota en la primera y segunda extracción es de 4(3).

Por lo tanto:

$$\text{Probabilidad (obtener dos sotas sin reemplazamiento)} = \frac{4(3)}{36(35)} = \frac{12}{1\,260} = \frac{1}{105}$$

Obtuvimos los mismos resultados.

Problema 41

Una bolsa contiene 5 canicas rojas, 4 azules y 6 verdes. Si se extrae en forma aleatoria una canica de la bolsa, calcula la probabilidad de que se extraigan en orden una roja, una azul y una verde si:

a) Las canicas se reemplazan

b) No se reemplazan

Solución:

A: canica roja en la primera extracción

B: canica azul en la segunda

C: canica verde en la tercera

a) Las canicas se reemplazan

Los sucesos son independientes y aplicamos:

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

$$= \frac{5}{15}\left(\frac{4}{15}\right)\left(\frac{6}{15}\right) = \frac{120}{3\,375}$$

$$= \frac{8}{225}$$

b) Las canicas no se reemplazan

Los sucesos son dependientes y aplicamos:

$$P(A \cap B \cap C) = P(A) \cdot P(B \mid A) \cdot P(C \mid A \cap B)$$

$$= \frac{5}{15}\left(\frac{4}{14}\right)\left(\frac{6}{13}\right) = \frac{120}{2\,730}$$

$$= \frac{4}{91}$$

Problema 42

Se lanza un dado dos veces sobre una mesa. Determina la probabilidad de que en la cara superior del dado salga al menos un 3 en alguno de los dos lanzamientos.

Solución:

A: un 3 en el primer lanzamiento

B: un 3 en el segundo lanzamiento

$A \cup B$: un 3 en el primer lanzamiento o un 3 en el segundo, o salgan dos 3.

Queremos obtener $P(A \cup B)$ y como los sucesos no son mutuamente excluyentes y son independientes, entonces:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= P(A) + P(B) - P(A) \cdot P(B)$$

$$= \frac{1}{6} + \frac{1}{6} - \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{2}{6} - \frac{1}{36} = \frac{12-1}{36} = \frac{11}{36}$$

Este problema también lo podemos resolver en la forma siguiente:

Como:

$P(\text{al menos un 3}) + P(\text{ningún 3}) = 1$

Despejando:

$$P(\text{al menos un 3}) = 1 - P(\text{ningún 3})$$

$$= 1 - P(A' \cap B')$$

$$= 1 - P(A') \cdot P(B')$$

$$= 1 - \left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = 1 - \frac{25}{36} = \frac{11}{36}$$

Problema 43

Si se tira un dado sobre una mesa y se lanza una moneda al aire, calcula la probabilidad de que salga, a la vez, un número menor que 3 y la cara de la moneda.

Solución:

A: Un número sea menor que 3

B: Caiga la cara de la moneda

La probabilidad de que un número sea menor que 3 es $P(A)=\frac{2}{6}=\frac{1}{3}$

La probabilidad de que caiga la cara de la moneda es $P(B)=\frac{1}{2}$

Como los sucesos son independientes, entonces:

$$P(A \text{ y } B) = P(A) \cdot P(B)$$
$$= \frac{1}{3}\left(\frac{1}{2}\right)$$
$$= \frac{1}{6}$$

Problema 44

En la oficina del subdirector de una escuela hay 12 calculadoras, algunas son manuales (M), otras eléctricas (E); además, algunas de ellas son nuevas (N) y otras usadas (U). Observa la siguiente tabla:

	M	E	
N	2	3	5
U	2	5	7
	4	8	12

a) Una persona entra a la oficina y elige aleatoriamente una calculadora y observa que es nueva. ¿Cuál es la probabilidad de que sea manual?

b) Si la persona escoge al azar una calculadora usada, ¿cuál es la probabilidad de que sea eléctrica?

c) ¿Cuál es la probabilidad de que la persona escoja una calculadora manual y usada?

Solución:

M: calculadora manual

E: calculadora eléctrica

N: calculadora nueva

U: calculadora usada

a) Queremos calcular $P(M \mid N)$

$$P(M \mid N) = \frac{P(M \cap N)}{P(N)} = \frac{\frac{2}{12}}{\frac{5}{12}} = \frac{24}{60}$$

$$= \frac{2}{5}$$

b) Queremos calcular $P(E \mid U)$

$$P(E \mid U) = \frac{P(E \cap U)}{P(U)} = \frac{\frac{5}{12}}{\frac{7}{12}} = \frac{60}{84}$$

$$= \frac{5}{7}$$

c) Queremos calcular $P(E' \cap U)$

$$P(E' \cap U) = P(U) \cdot P(E' \mid U) = \frac{7}{12}\left(\frac{2}{7}\right) = \frac{14}{84}$$

$$= \frac{1}{6}$$

o también:

$$P(E' \cap U) = P(E') \cdot P(U \mid E') = \frac{4}{12}\left(\frac{2}{4}\right) = \frac{8}{48}$$

$$= \frac{1}{6}$$

Para calcular $P(E' \mid U)$, el espacio muestral se reduce a 7 elementos, de los cuales 2 son favorables al suceso E'.

Asimismo, para calcular $P(U \mid E')$ el espacio muestral se reduce a 4 elementos, de los cuales 2 son favorables para el suceso U.

Problema 45

KILOMEX es una fábrica que produce transistores. En un lote de 100 piezas, 80 artículos no presentan algún defecto, mientras que 20 no cumplen con las normas de calidad. Si se elijen al azar 2 transistores sin sustitución, ¿cuál es la probabilidad de que ambos artículos sean defectuosos?

Solución:

A: el primer transistor es defectuoso

B: el segundo transistor es defectuoso

Queremos calcular:

$$P(A \text{ y } B) = P(A) \cdot P(B \mid A)$$
$$= \frac{20}{100}\left(\frac{19}{99}\right) = \frac{380}{9\,900}$$
$$= 0.038$$
$$= 3.8\%$$

Resumen

1. Para sucesos mutuamente excluyentes se tiene $A \cap B = \varnothing$ y se aplica la siguiente relación:

$$P(A \text{ o } B) = P(A) + P(B)$$

2. Cuando los sucesos A y B no son mutuamente excluyentes, se tiene $A \cap B \neq \varnothing$ y se aplica la siguiente relación:

$$P(A \text{ o } B) = P(A) + P(B) - P(A \cap B)$$

Recuerda que para $P(A \text{ o } B) = P(A \cup B)$ usarás la notación que indique tu profesor.

3. Cuando los sucesos A y B son independientes, se aplica la siguiente relación:

$$P(A \text{ y } B) = P(A) \cdot P(B)$$

Recuerda que para $P(A \text{ y } B) = P(A \cap B)$ usarás la notación que indique tu profesor.

4. Cuando los sucesos A y B son dependientes, se aplica la relación de la probabilidad condicional:

$$P(A \text{ y } B) = P(A) \cdot P(B \mid A)$$

5. La probabilidad condicional de B habiendo sucedido A es la relación:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \text{ con } P(A) > 0$$

Las probabilidades de la probabilidad condicional son:

a) Para sucesos dependientes:

$$P(A \cap B) = P(A) \cdot P(B \mid A)$$

Además:

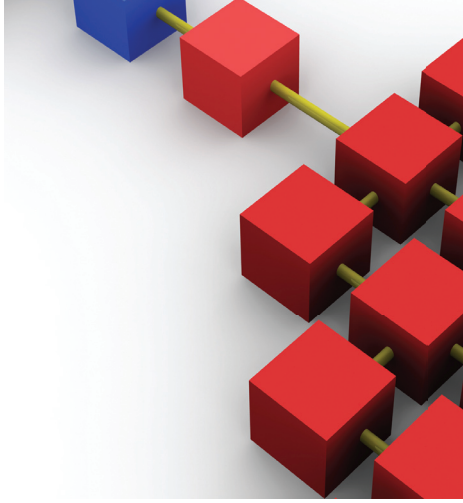
$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2)$$

b) Para sucesos independientes, tales que $P(A) > 0$ y $P(B) > 0$

se aplica la relación $P(A \cap B) = P(A) \cdot P(B)$.

Además, si A y B son sucesos también independientes, se aplica la relación

$P(A \cap B) = P(A) + P(B) - P(A) \cdot P(B)$



Capítulo 7

Análisis combinatorio y probabilidad

Procesos estocásticos

Regla de Bayes

Conceptos clave

Permutación
Proceso estocástico
Regla de Bayes

Análisis combinatorio y probabilidad

En algunos casos resulta fácil obtener los puntos muestrales de un espacio y por su número se puede realizar la enumeración directa necesaria para obtener las probabilidades. Sin embargo, hay problemas en los que no es posible realizar la cuenta directa. En estos casos, se emplea el cálculo combinatorio que, como sabes, es una forma especial de conteo.

Problema 1

a) Raúl y nueve amigos se quieren formar en fila. ¿Cuál es la probabilidad de que Raúl quede al frente?

Solución:

Para calcular el tamaño del espacio muestral razonamos así: cualquiera de los diez alumnos puede ocupar el primer lugar, pero para el segundo sitio sólo hay nueve elecciones posibles puesto que hay un alumno colocado al frente. El tercer sitio se puede escoger de entre ocho alumnos, y así sucesivamente hasta el último lugar que, una vez colocados los anteriores, se ocupa con un alumno.

Dado que la **permutación** se define como las diferentes maneras en que se pueden ordenar los elementos, se tiene que:

${}_{10}P_{10} = 10\ (9)\ (8)\ (7)\ (6)\ (5)\ (4)\ (3)\ (2)\ (1) = 10! = 3\ 628\ 800$ y para que Raúl quede al frente de sus nueve compañeros se puede ordenar de ${}_{9}P_{9}$ maneras:

$$ {}_{9}P_{9} = 9\ (8)\ (7)\ (6)\ (5)\ (4)\ (3)\ (2)\ (1) = 9! = 362\ 880 $$

Entonces, la probabilidad de que Raúl quede al frente es:

$$ P(R) = \frac{{}_{9}P_{9}}{{}_{10}P_{10}} = \frac{9(8)(7)(6)(5)(4)(3)(2)(1)}{10(9)(8)(7)(6)(5)(4)(3)(2)(1)} = \frac{1}{10} = 0.1 $$

b) Ahora queremos saber cuál es la probabilidad de que los primeros cuatro lugares estén ocupados por Raúl, Antonio, Juan y Pedro, en ese orden.

Se tiene ${}_6P_6$:

$${}_6P_6 = 6\ (5)\ (4)\ (3)\ (2)\ (1) = 720$$

La probabilidad de que Raúl, Antonio, Juan y Pedro queden al frente es:

$$P(R, A, J, P) = \frac{{}_6P_6}{{}_{10}P_{10}} = \frac{720}{3\,628\,800} = 0.00019$$

Problema 2

Un niño está jugando con cinco soldados iguales, tres jinetes iguales y seis caballos iguales. ¿Cuál es la probabilidad de que en los cinco primeros lugares queden los cinco soldados?

Solución:

Primero calculamos el tamaño del espacio muestral, calculando de cuántas maneras diferentes se pueden ordenar los juguetes.

En la fórmula $P = \dfrac{n!}{n_1!n_2!n_3!\ldots n_n!}$ sustituimos con:

$n = (5 + 3 + 6) = 14$

$n_1 = 5$

$n_2 = 3$

$n_3 = 6$

$$P_{14} = \frac{14!}{5!3!6!} = \frac{14(13)(12)(11)(10)(9)(8)(7)(6)(5)(4)(3)(2)(1)}{5(4)(3)(2)(1)\ \ 3(2)(1)\ \ 6(5)(4)(3)(2)(1)}$$

$$= \frac{14(13)(11)(3)(4)(7)}{1}$$

$$= 168\,168$$

Hay 168 168 maneras de ordenar los juguetes.

Si los soldados quedan al frente, los juguetes restantes se pueden ordenar así:

$$P = \frac{9!}{3!6!} = \frac{9(8)(7)(6)(5)(4)(3)(2)(1)}{3(2)(1)\ \ (6)(5)(4)(3)(2)(1)} = \frac{84}{1}$$

La probabilidad es:

$$P\ (\text{cinco soldados al frente}) = \frac{84}{168\,168} = 0.00049$$

Problema 3

En una oficina hay ocho personas que tienen que dividirse en dos equipos para realizar una tarea; un equipo debe tener seis integrantes y el otro dos.

Solución

Una vez seleccionadas dos personas para un equipo quedan determinadas las otras seis para integrar el otro equipo. Dado que no nos interesa el orden que las personas tengan en cada equipo, podemos entender el problema como si en ocho elementos escogiéramos de dos en dos.

Solución:

En la fórmula $\binom{n}{r} = {}_nC_r = \dfrac{n!}{(n-r)!\ r!}$ sustituimos con:

$n = 8$

$r = 2$

$$\binom{8}{2} = {}_8C_2 = \frac{8!}{(8-2)!2!} = \frac{8!}{6!2!} = \frac{\overset{4}{\cancel{8}}(7)(\cancel{6})}{\cancel{6}!\cancel{2}(1)} = 28$$

Hay 28 maneras de formar los dos equipos.

Para que María y Saúl queden juntos en el equipo 2, sucede sólo una vez de las 28, por lo tanto:

$$P(M,S) = \frac{1}{28} = 0.0357$$

Problema 4

Se organiza un comité de cinco personas en el que debe haber dos arquitectos, de siete que hay en la compañía, y tres ingenieros, de los diez que trabajan también en esa compañía. ¿Cuál es la probabilidad de que en el comité queden Sergio y Rogelio, que son arquitectos, y Eduardo, que es ingeniero?

Solución:

Calculamos el tamaño del espacio muestral con la siguiente fórmula:

$$\binom{n}{r} = {}_nC_r = \frac{n!}{(n-r)!\ r!}$$

Obtenemos las combinaciones para los arquitectos al sustituir con:

$n = 7$

$r = 2$

$$\binom{7}{2} = {}_7C_2 = \frac{7!}{(7-2)!2!} = \frac{7!}{5!2!} = \frac{7(\overset{3}{\cancel{6}})(\cancel{5}!)}{\cancel{5}!(\cancel{2})(1)} = 21$$

Para las combinaciones de los ingenieros sustituimos con:

$n = 10$

$r = 3$

$$\binom{10}{3} = {}_{10}C_3 = \frac{10!}{(10-3)!3!} = \frac{10!}{7!3!} = \frac{10(\overset{3}{\cancel{9}})(\overset{4}{\cancel{8}})(\cancel{7!})}{\cancel{7!}(\cancel{3})(\cancel{2})(1)} = 120$$

Por el principio multiplicativo del conteo obtenemos:

$${}_7C_2 \cdot {}_{10}C_3 = \frac{7!}{(7-2)!2!} \cdot \frac{10!}{(10-3)!3!} = 21(120) = 2\,520$$

Hay 2 520 maneras diferentes de integrar el comité.

El comité en el que Sergio y Rogelio son los elegidos como arquitectos es uno de los 21 que se pueden formar.

Para que Eduardo sea elegido, los otros dos ingenieros se eligen de los nueve restantes:

$$\binom{9}{2} = {}_9C_2 = \frac{9!}{(9-2)!(2!)} = \frac{9(8)(\cancel{7!})}{(\cancel{7!})(2!)} = 36$$

El número de comités en los que Sergio, Rogelio y Eduardo están es de 36 y la probabilidad de que sean elegidos es:

$$P = \frac{36}{2\,520} = 0.0142$$

Si se busca que en el comité quede un ingeniero en particular, ¿cuál sería la probabilidad de que esto suceda?

Solución:

Los arquitectos quedan igual: ${}_7C_2 = 21$

Los dos ingenieros que faltan se toman de nueve y pueden elegirse así:

$$\binom{9}{2} = {}_9C_2 = \frac{9!}{(9-2)!2!} = \frac{9!}{7!2!} = \frac{9(8)(\cancel{7!})}{\cancel{7!}\,2!} = 36$$

Por el principio multiplicativo del conteo, se tiene:

$${}_7C_2 \cdot {}_9C_2 = 21(36) = 756$$

Para un comité donde se incluye un ingeniero en particular la probabilidad es:

$$\frac{756}{2\,520} = 0.3$$

Problema 5

En una escuela cuatro profesores pueden ser seleccionados para participar en un proyecto de actualización; el director no sabe cuántos profesores de los cuatro van a ser requeridos: puede ser uno, dos, tres o los cuatro. ¿Cuál es la probabilidad de que sean seleccionados sólo dos profesores?

Solución:

Primero calculamos el tamaño del espacio muestral:

- Si un profesor es seleccionado ${}_4C_1$
- Si son dos ${}_4C_2$
- Si son tres ${}_4C_3$
- Si son cuatro ${}_4C_4$

Por el principio aditivo del conteo, tenemos:

$$ {}_4C_1 + {}_4C_2 + {}_4C_3 + {}_4C_4 $$

$$ = \frac{4!}{(4-1)!1!} + \frac{4!}{(4-2)!2!} + \frac{4!}{(4-3)!3!} + \frac{4!}{(4-4)!4!} = 4+6+4+1=15 $$

Hay 15 maneras de seleccionar a los profesores.

Como tenemos ${}_4C_2$ maneras de que se seleccionen sólo dos profesores, la probabilidad de que esto suceda es:

$$ P = \frac{{}_4C_2}{15} = \frac{6}{15} = 0.4 $$

Problema 6

En una mesa hay cinco verduras diferentes. ¿Cuál es la probabilidad de que una persona coma tres verduras diferentes?

Solución:

Primero calculamos el tamaño del espacio muestral:

Una persona coma una verdura ${}_5C_1$

Una persona coma dos verduras diferentes ${}_5C_2$

Una persona coma tres verduras diferentes ${}_5C_3$

Una persona coma cuatro verduras diferentes ${}_5C_4$

Una persona coma todas las verduras disponibles ${}_5C_5$

Por el principio aditivo del conteo, tenemos:

$$ {}_5C_1 + {}_5C_2 + {}_5C_3 + {}_5C_4 + {}_5C_5 $$

$$= \frac{5!}{(5-1)!1!} + \frac{5!}{(5-2)!2!} + \frac{5!}{(5-3)!3!} + \frac{5!}{(5-4)!4!} + \frac{5!}{(5-5)!5!}$$

$$= 5 + 10 + 10 + 5 + 1 = 31$$

La probabilidad de que una persona coma tres verduras es:

$$P = \frac{{}_5C_3}{31} = \frac{10}{31} = 0.3225$$

Problema 7

En un librero hay ocho libros de física y cinco de química. Determina la probabilidad de que tres libros en particular de física queden juntos.

Solución:

$8 + 5 = 13$ libros pueden ordenarse entre sí de:

${}_{13}P_{13} = 13!$ formas

Si los tres libros deseados de física los consideramos como uno solo, tenemos $13 - 3 = 10$; $10 + 1 = 11$ libros que podemos ordenar entre sí de ${}_{11}P_{11} = 11!$ formas.

Por lo tanto, la probabilidad P es:

$$P = \frac{11!3!}{13!} = \frac{11(10)(9)(8)(7)(6)(5)(4)(3)(2)(1)(3)(2)(1)}{13(12)(11)(10)(9)(8)(7)(6)(5)(4)(3)(2)(1)} = \frac{6}{156}$$

$$P = \frac{6}{156} = \frac{1}{26}$$

Hay $\frac{1}{26}$ de probabilidad de que los tres libros deseados de física queden juntos.

Problema 8

Diez alumnos se van a formar en fila. ¿Cuál es la posibilidad de que Pedro y Juan queden juntos?

Solución:

Los diez alumnos se pueden ordenar de ${}_{10}P_{10} = 10!$ formas.

Dado que buscamos que Pedro y Juan queden juntos, se considera esto como un solo lugar. Se tiene ahora $10 - 2 = 8$; $8 + 1 = 9$ personas que se pueden ordenar de ${}_9P_9 = 9!$ maneras diferentes.

Como Pedro y Juan se pueden ordenar entre sí de ${}_2P_2 = 2!$ formas.

Se tiene que la probabilidad de que queden juntos es de

$$P = \frac{9!2!}{10!} = \frac{9!2!}{10(9!)} = \frac{2(1)}{10} = \frac{1}{5} = 0.2$$

$$= 20\%$$

Problema 9

Una bolsa contiene siete canicas rojas, tres blancas y diez verdes. Se extraen tres canicas al azar sin reemplazo. Calcula la probabilidad de que:

a) Las tres canicas sean rojas.

b) Las tres canicas sean blancas.

c) Dos canicas sean rojas y una blanca.

d) Al menos una sea blanca.

e) Se extraiga una canica de cada color.

Solución:

$n(A)$ roja = 7

$n(B)$ blanca = 3

$n(C)$ verde = 10

Total de canicas: 20

a) $P(A \cap B \cap C) = P(A) \cdot P(B \mid A) \cdot P(C \mid A \cap B)$

$$= \left(\frac{7}{20}\right)\left(\frac{6}{19}\right)\left(\frac{5}{18}\right) = \frac{210}{6\,840}$$

$$= \frac{7}{228} = 0.0307 \quad \text{(se tomaron 4 cifras decimales)}$$

$$= 3.07\%$$

Ahora resolvemos el mismo problema, inciso **a)** aplicando las combinaciones.

$$\text{Probabilidad (tres canicas rojas)} = \frac{\text{Número de grupos de tres canicas entre siete rojas}}{\text{Número de grupos de tres canicas entre veinte}}$$

$$= \frac{{}_7C_3}{{}_{20}C_3} = \frac{\dfrac{7!}{(7-3)!3!}}{\dfrac{20!}{(20-3)!3!}} = \frac{\dfrac{7!}{4!3!}}{\dfrac{20!}{17!3!}} = \frac{17!7!}{20!4!} = \frac{7(6)(5)}{20(19)(18)}$$

$$= \frac{210}{6\,840} = \frac{7}{228} = 0.0307$$

$$= 3.07\%$$

Obtuvimos el mismo resultado que el anterior.

b) Probabilidad (tres canicas blancas) $= \dfrac{{}_3C_3}{{}_{20}C_3} = \dfrac{\dfrac{3!}{(3-3)!3!}}{\dfrac{20!}{(20-3)!3!}}$

$$= \frac{\dfrac{3!}{0!3!}}{\dfrac{20!}{17!3!}} = \frac{17!3!}{1(20!)}$$

$$= \frac{3(2)}{18(19)(20)} = \frac{6}{6\,840}$$

$$= 0.0008 \text{ se tomaron 4 cifras}$$

$$= 0.08\%$$

c) Probabilidad (dos canicas rojas y una blanca) $= \dfrac{\left({}_7C_2\right)\left({}_3C_1\right)}{{}_{20}C_3}$

$$= \frac{\dfrac{7!}{(7-2)!2!} \cdot \dfrac{3!}{(3-1)!1!}}{\dfrac{20!}{(20-3)!3!}}$$

$$= \frac{\dfrac{7!3!}{5!2!2!1}}{\dfrac{20!}{17!3!}} = \frac{7!3!17!3!}{5!2!2!20!}$$

$$= \frac{7(6)(3)(3)}{20(19)(18)}$$

$$= \frac{378}{6\,840} = \frac{126}{2\,280}$$

$$= 0.0552$$

$$= 5.52\%$$

d) Probabilidad (al menos una canica blanca)
Inicialmente calculamos para ninguna canica blanca

Probabilidad (ninguna canica blanca) $= \dfrac{{}_{17}C_3}{{}_{20}C_3} = \dfrac{\dfrac{17!}{(17-3)!3!}}{\dfrac{20!}{(20-3)!3!}} = \dfrac{\dfrac{17!}{14!3!}}{\dfrac{20!}{17!3!}}$

$$= \frac{17!17!3!}{20!14!3!}$$

$$= \frac{17(16)(15)}{20(19)(18)} = \frac{4\,080}{6\,840}$$

$$= \frac{34}{57}$$

Probabilidad (al menos una canica blanca)
$$= 1 - \frac{34}{57} = \frac{57}{57} - \frac{34}{57}$$

$$= \frac{23}{57} = 0.4035 \text{ (se tomaron cuatro cifras decimales)}$$

$$= 40.35\%$$

e) Probabilidad (una canica de cada color)
$$= \frac{\left({}_7C_1\right)\left({}_3C_1\right)\left({}_{10}C_1\right)}{{}_{20}C_3} =$$

$$= \frac{\frac{7!}{(7-1)!1!} \cdot \frac{3!}{(3-1)!1!} \cdot \frac{10!}{(10-1)!1!}}{\frac{20!}{(20-3)!3!}}$$

$$= \frac{7!3!10!17!3!}{6!2!9!20!}$$

Recuerda que el factorial de 1! = 1 y el uno es el elemento neutro de la multiplicación.

$$= \frac{7(3)(2)(10)(3)(2)}{2(20)(19)(18)}$$

$$= \frac{7}{38} = 0.1842 \text{ (se tomaron cuatro cifras decimales)}$$

$$= 18.42\%$$

Procesos estocásticos

Una sucesión finita de experimentos en los cuales cada experimento tiene un número finito de resultados con probabilidades dadas se conoce como **proceso estocástico**.

La sucesión de experimentos puede realizarse en forma simultánea o sucesiva de dos o más experimentos aleatorios diferentes o idénticos.

Los experimentos son independientes, ya que los resultados que se llevan a cabo inicialmente, o de manera simultánea, no afectan los resultados de ninguno de los otros experimentos que pudieran intervenir.

En consecuencia, los puntos muestrales que van a combinarse se pueden considerar como eventos independientes y su verificación simultánea puede interpretarse como una intersección, por lo que la probabilidad de su ocurrencia simultánea se puede obtener con base en el principio multiplicativo de probabilidades y procedemos así:

- Se desarrolla el espacio muestral del proceso auxiliándose de un diagrama de árbol que, a diferencia de los que se citaron en el tema de conjuntos, deberán incluir en la parte media de cada rama la probabilidad de que en su espacio muestra tiene asociada los puntos muestra simples, mediante algunas anotaciones añadidas en el diagrama que señalen de alguna manera la forma en que se va a desarrollar el proceso.
- Los puntos muestra se obtienen en la forma ya estudiada y sus probabilidades corresponderán al producto de todas las probabilidades simples que se encuentren en el recorrido correspondiente.

Problema 10

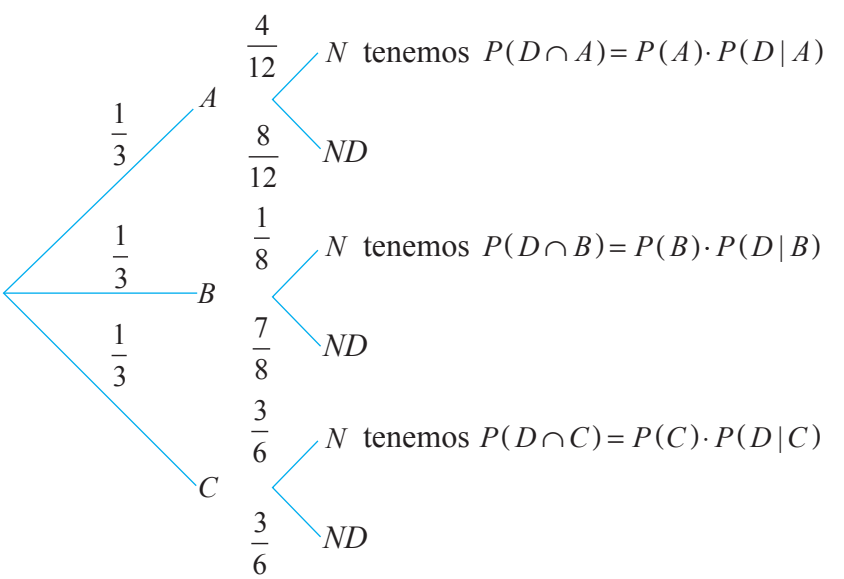
En una tienda de material de carburadores hay tres cajas. La primera caja contiene doce carburadores, de los cuales cuatro están defectuosos; la segunda caja contiene ocho carburadores con uno defectuoso; y la tercera caja contiene seis carburadores con tres defectuosos. Si se toma aleatoriamente una de las tres cajas y luego se toma al azar un carburador, determina la probabilidad de que el carburador esté defectuoso.

Solución:

(A) Primera caja

(B) Segunda caja

(C) Tercera caja



Probabilidad (artículo defectuoso) $= P(A)\cdot P(D\mid A)+P(B)\cdot P(D\mid B)+P(C)\cdot P(D\mid C)$

$$P(D)=\frac{1}{3}\left(\frac{4}{12}\right)+\frac{1}{3}\left(\frac{1}{8}\right)+\frac{1}{3}\left(\frac{3}{6}\right)=\frac{4}{36}+\frac{1}{24}+\frac{3}{18}$$

$$=\frac{8+3+12}{72}=\frac{23}{72}=0.3194 \text{ (se tomaron 4 cifras decimales)}$$

$$=31.94\%$$

Problema 11

Un pequeño restaurante ofrece comida corrida. Los comensales pueden escoger entre tres sopas: arroz, espagueti y sopa de verduras; cuatro guisados: pollo, bistec, enchiladas y tortas de papa; y dos postres: arroz con leche y gelatina.

¿Cuál es la probabilidad de que un comensal escoja arroz, bistec y gelatina?

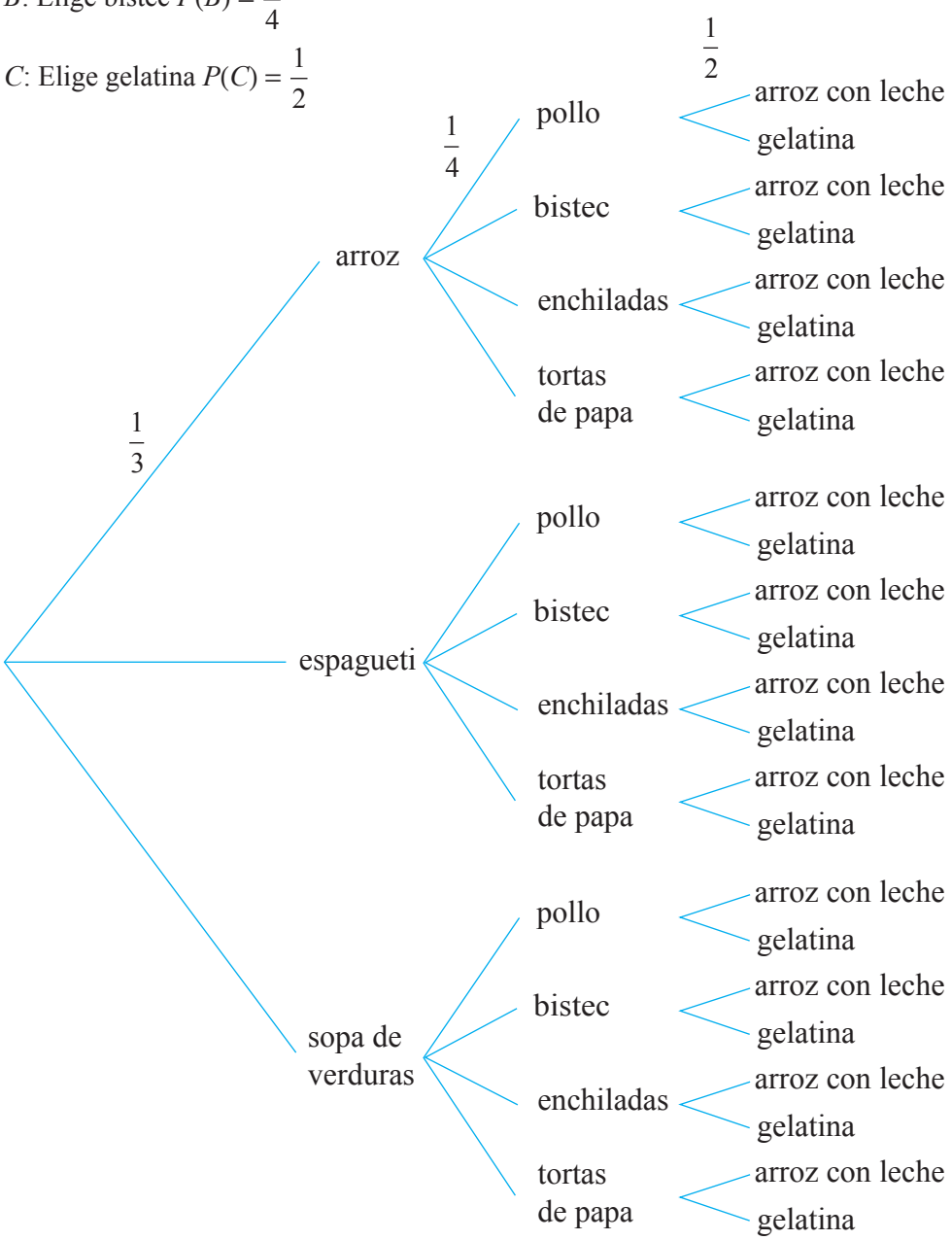
Solución:

Sean:

A: Elige arroz $P(A) = \frac{1}{3}$

B: Elige bistec $P(B) = \frac{1}{4}$

C: Elige gelatina $P(C) = \frac{1}{2}$



$$P(A\cap B\cap C)=P(A)\cdot P(B\mid A)\cdot P(C\mid A\cap B)$$
$$=\frac{1}{3}\left(\frac{1}{4}\right)\left(\frac{1}{2}\right)=\frac{1}{24}=0.0416 \text{ (se tomaron 4 cifras decimales)}$$
$$=4.16\%$$

Problema 12

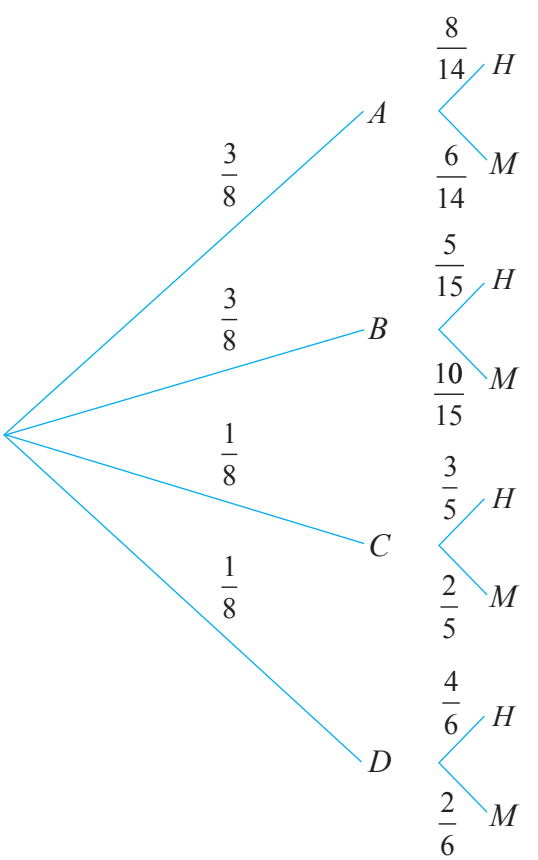
Se va a sortear una beca entre los mejores promedios de cuatro escuelas. Las escuelas A y B tienen $\frac{3}{8}$ de probabilidad de obtener la beca, mientras que las escuelas C y D tienen $\frac{1}{8}$.

- La escuela A tiene ocho hombres y seis mujeres entre sus mejores promedios.
- La escuela B tiene cinco hombres y diez mujeres.
- La escuela C tiene tres hombres y dos mujeres.
- La escuela D tiene cuatro hombres y dos mujeres.

Si Ana está en la escuela C, ¿qué probabilidad tiene de ser elegida como becaria?

Solución:

$$P(C\cap M)=P(C)\cdot P(M\mid C)$$
$$=\frac{1}{8}\left(\frac{1}{5}\right)=\frac{1}{40}=0.25$$
$$=2.5\%$$



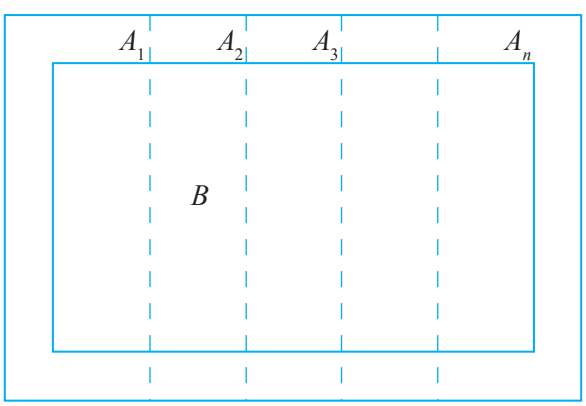
Regla de Bayes

La regla de Bayes simplifica el cálculo de las probabilidades con condicionales. Permite calcular la probabilidad de que ocurra el suceso B si se sabe que ya ocurrió el suceso A. Se expresa $P(B \mid A)$ y para ello se necesita conocer la probabilidad como frecuencia relativa de que ocurra el suceso A, o sea $P(A)$.
La probabilidad de que ocurra el suceso B es $P(B)$.
La probabilidad como frecuencia relativa de que ocurra el suceso B es $P(B)$.
La probabilidad de que ocurra el suceso A si sabemos que ya ocurrió el suceso B es $P(A \mid B)$.

Razonamiento para obtener la regla de Bayes

Si los sucesos A_1, A_2, A_3… A_n forman una partición de un espacio muestral S, entonces los sucesos citados son mutuamente excluyentes y su unión es S. La suma de probabilidades es igual a la unidad o al 100%.

$$A_1 \cup A_2 \cup A_3 \cup \ldots A_n = S$$
$$A_1 \cap A_2 \cap A_3 \cap \ldots A_n = \phi$$



Si B es otro suceso dentro del espacio muestral tenemos:

$$B \subset S$$

$$B = S \cap B$$

$$B = (A_1 \cup A_2 \cup A_3 \cup \ldots A_n) \cap B$$

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \ldots \cup (A_n \cap B)$$

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots P(A_n \cap B) \quad \textbf{(1)}$$

$$P(A_i \cap B) = P\left(A_i\right) \cdot P(B \mid A_i) \quad \textbf{(2)}$$

Sustituyendo **(2)** en **(1)** para $i = 1, 2, \ldots n$

$$P(B) = P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2) + \ldots + P(A_n) \cdot P(B \mid A_n) \quad \textbf{(3)}$$

Por otro lado,

$$P(A_i \mid B) = \frac{P(B \cap A_i)}{P(B)} \quad \textbf{(4)}$$

Sustituyendo **(2)** y **(3)** en **(4)**

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{P(A_i) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2) + \ldots + P(A_n) \cdot P(B \mid A_n)}$$

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{\sum_{j=i}^{n} P(A_j) \cdot P(B \mid A_j)}$$

$j = 1, 2, \ldots n$

La regla de Bayes en su forma más sencilla queda así:

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)}$$

A la fórmula de Bayes también se le conoce como la *probabilidad de las causas*, puesto que las A_1 son una partición del espacio muestral S, uno y sólo uno de los sucesos A_i ocurre; o sea que uno de los sucesos A_i debe ocurrir y solamente uno.

La fórmula anterior nos da la probabilidad de que un suceso A_i particular ocurra (donde A_i es una causa) dado que B ha ocurrido.

Problema 13

En una maquiladora de circuitos para computadoras, un grupo de trabajadores A produce 68% de la producción total de la maquiladora, mientras que el grupo B produce 32% del total de la producción.

Del grupo A, 2% de su producción es defectuosa, en tanto que la producción del grupo B presenta 5% de artículos defectuosos. Calcula cuál es la probabilidad de que al revisar al azar un circuito resulte defectuoso.

a) ¿Cuál es la probabilidad de que al revisar aleatoriamente un circuito defectuoso provenga del grupo B?

Solución:

a) La probabilidad de revisar un circuito defectuoso es el promedio del porcentaje de artículos defectuosos que produce cada grupo, ponderado por un porcentaje de participación en la producción total de la maquiladora.

$$\text{Probabilidad de circuitos defectuosos} = (0.02)(0.68)+(0.05)(0.32)$$
$$= 0.0136 + 0.016$$
$$= 0.0296$$
$$= 2.96\%$$

b) Ahora aplicaremos la fórmula de Bayes para calcular la probabilidad de que al revisar aleatoriamente un circuito defectuoso éste provenga del grupo B:

$P(A)$ como la probabilidad de frecuencia relativa de obtener un circuito defectuoso, que es de 2.96%.

$P(B)$ probabilidad como frecuencia relativa de obtener un circuito del grupo B, que es de 32%.

$P(A \mid B)$ probabilidad condicional de obtener un circuito defectuoso, que si procede del grupo B, es de 5%.

Sustituimos en:

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)}$$
$$= \frac{(0.05)(0.32)}{0.0296}$$
$$= 0.5405 \text{ (se tomaron 4 cifras decimales)}$$
$$= 54.05\%$$

Problema 14

Una fábrica productora de artículos metálicos cuenta con tres sucursales, las cuales producen 40, 35 y 25% del total de la producción, respectivamente. Sin embargo, en cada sucursal se presentan los siguientes porcentajes de artículos defectuosos: 4, 6 y 8%, respectivamente. Si se elige aleatoriamente un artículo, calcula cuál es la probabilidad:

a) De que el artículo no sea defectuoso.

b) Si el artículo resultó defectuoso, cuál es la probabilidad de que proceda de la primera sucursal.

c) Si el artículo no resultó defectuoso, cuál es la probabilidad de que proceda de la segunda sucursal.

Solución:

A_1: El producto es de la primera sucursal.

A_2: El producto es de la segunda sucursal.

A_3: El producto es de la tercera sucursal.

B: El producto es defectuoso.

a) Probabilidad de que el artículo no sea defectuoso.

$$P(A_1) = 40\% = \frac{2}{5}$$

$$P(A_2) = 35\% = \frac{7}{20}$$

$$P(A_3) = 25\% = \frac{1}{4}$$

Además:

$$P(B \mid A_1) = 4\% = \frac{1}{25}$$

$$P(B \mid A_2) = 6\% = \frac{3}{50}$$

$$P(B \mid A_3) = 8\% = \frac{2}{25}$$

$$P(B) = P(B \mid A_1) \cdot P(A_1) + P(B \mid A_2) \cdot P(A_2) + P(B \mid A_3) \cdot P(A_3)$$

$$P(B) = \frac{1}{25}\left(\frac{2}{5}\right) + \frac{3}{50}\left(\frac{7}{20}\right) + \frac{2}{25}\left(\frac{1}{4}\right)$$

$$P(B) = \frac{2}{125} + \frac{21}{1\,000} + \frac{2}{100} = \frac{16 + 21 + 20}{1\,000} = \frac{57}{1\,000}$$

$$P(B') = 1 - \frac{57}{1\,000} = \frac{1000}{1\,000} - \frac{57}{1\,000}$$

$$= \frac{943}{1\,000} = 0.943$$

$$= 94.3\%$$

b) Si el artículo resultó defectuoso, cuál es la probabilidad de que proceda de la primera sucursal.

$$P(A_1 \mid B) = \frac{P(B \mid A_1) \cdot P(A_1)}{P(B)} = \frac{\frac{1}{25}\left(\frac{2}{5}\right)}{\frac{57}{1\,000}} = \frac{\frac{2}{125}}{\frac{57}{1\,000}} = \frac{2\,000}{7\,125}$$

$$= \frac{16}{57} = 0.2807 \quad \text{(se tomaron 4 cifras decimales)}$$

$$= 28.07\%$$

c) Si el artículo no resultó defectuoso, cuál es la probabilidad de que proceda de la segunda sucursal.

$$P(A_2 \mid B') = \frac{P(A_2 \cap B')}{P(B')} = \frac{P(A_2) - P(A_2 \cap B)}{P(B')}$$

$$= \frac{P(A_2) - P(B \mid A_2) \cdot P(A_2)}{P(B')}$$

$$= \frac{\frac{7}{20} - \left(\frac{3}{50}\right)\left(\frac{7}{20}\right)}{\frac{943}{1\,000}} = \frac{\frac{7}{20} - \frac{21}{1\,000}}{\frac{943}{1\,000}} = \frac{\frac{350-21}{\cancel{1000}}}{\frac{943}{\cancel{1000}}}$$

$$= \frac{329}{943} = 0.3488$$

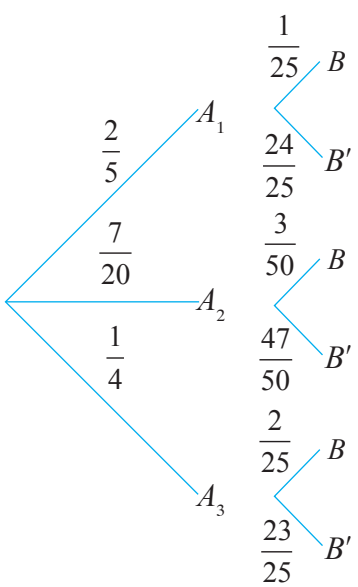
$$= 34.88\%$$

Los problemas que se resuelven con la regla de Bayes nos permiten analizar algunos problemas aplicando el diagrama de árbol al enfocarlos como procesos estocásticos. Para ello, el diagrama deberá tener tantas ramificaciones primarias como particiones existan; además, cada terminal primaria tendrá dos ramificaciones secundarias que representarán, respectivamente, los puntos muestrales que queden dentro del evento B.

Problema 15

Resolveremos nuevamente el problema anterior aplicando el diagrama de árbol.

Solución:



a) De que el artículo no sea defectuoso.

$$P(B') = \frac{2}{5}\left(\frac{24}{25}\right) + \frac{7}{20}\left(\frac{47}{50}\right) + \frac{1}{4}\left(\frac{23}{25}\right)$$

$$= \frac{48}{125} + \frac{329}{1\,000} + \frac{23}{100}$$

$$= \frac{384 + 329 + 230}{1\,000} = 0.943$$

$$= 94.3\%$$

b) Si el artículo resultó defectuoso, cuál es la probabilidad de que proceda de la primera sucursal.

$$P(A_1 \mid B) = \frac{\frac{2}{5}\left(\frac{1}{25}\right)}{\frac{57}{1\,000}} = \frac{\frac{2}{125}}{\frac{57}{1\,000}} = \frac{2\,000}{7\,125} = \frac{16}{57} = 0.2807$$
$$= 28.07\%$$

c) Si el artículo no resultó defectuoso, cuál es la probabilidad de que proceda de la sucursal A_2.

$$P(A_2 \mid B') = \frac{\frac{7}{20}\left(\frac{47}{50}\right)}{\frac{943}{1\,000}} = \frac{\frac{329}{1\,000}}{\frac{943}{1\,000}} = \frac{329}{943} = 0.3488$$
$$= 34.88\%$$

Obtuvimos los mismos resultados.

Problema 16

En una escuela primaria, 40% de los alumnos cursan el primer año, 25% el segundo año, 20% el tercer año y 15% el último año.

Los porcentajes de alumnos que asisten al taller de teatro son los siguientes:

100% los de primer año.

40% los de segundo año.

20% los de tercer año.

10% los del último año.

Si se escoge aleatoriamente un alumno que asiste al taller de teatro, ¿cuál es la probabilidad de que sea del segundo año?

Solución:

A_1: Cursa el primer año $P(A_1) = 40\% = \frac{2}{5}$

A_2: Cursa el segundo año $P(A_2) = 25\% = \frac{1}{4}$

A_3: Cursa el tercer año $P(A_3) = 20\% = \frac{1}{5}$

A_4: Cursa el último año $P(A_4) = 15\% = \frac{3}{20}$

B: Asiste al taller de teatro

B_1: Cursa el primer año y asiste al taller de teatro

$$P(B_1) = 100\%\left(\frac{2}{5}\right) = P(A_1 \cap B)$$

B_2: Cursa el segundo año y asiste al taller de teatro

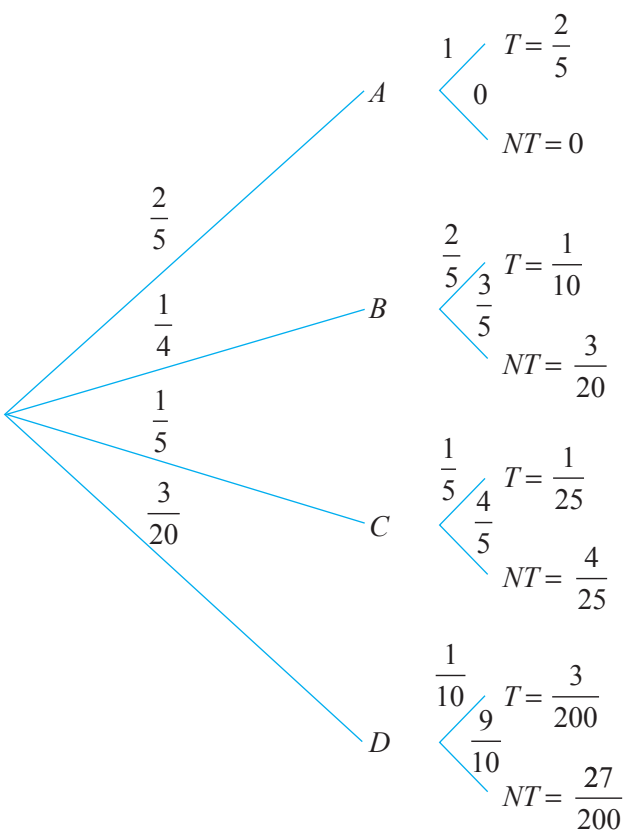
$$P(B_2) = 40\%\left(\frac{1}{4}\right) = P(A_2 \cap B)$$

B_3: Cursa el tercer año y asiste al taller de teatro

$$P(B_3) = 20\%\left(\frac{1}{5}\right) = P(A_3 \cap B)$$

B_4: Cursa el cuarto año y asiste al taller de teatro

$$P(B_4) = 10\%\left(\frac{3}{20}\right) = P(A_4 \cap B)$$



$= P$ (asiste al taller de teatro y cursa el segundo año) $= P(A_2 \cap B)$

$$= \left(\frac{1}{4}\right)\left(\frac{2}{5}\right) = \frac{2}{20}$$

$$= \frac{1}{10}$$

A partir de la fórmula de probabilidad condicional, se sabe que la probabilidad de que el alumno curse el segundo año (A_2), sabiendo que estudia teatro (B), es:

$$P(A_2 \mid B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{\frac{1}{10}}{P(B)}$$

Para tener un alumno que estudie teatro, éste puede ser de cualquiera de los cuatro años:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + P(A_4 \cap B)$$
$$= \frac{2}{5} + \frac{1}{10} + \frac{1}{25} + \frac{3}{200} = \frac{80 + 20 + 8 + 3}{200} = \frac{111}{200}$$

Por lo tanto,

$$P(A_2 \mid B) = \frac{\frac{1}{10}}{\frac{111}{200}} = \frac{200}{1\,110} = 0.18$$

$$= 18\%$$

Problema 17

El problema anterior también se puede resolver con la regla de Bayes, así:

$$P(A_2 \mid B) = \frac{P(A_2) \cdot P(B \mid A_2)}{P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2) + P(A_3) \cdot P(B \mid A_3) + P(A_4) \cdot P(B \mid A_4)}$$

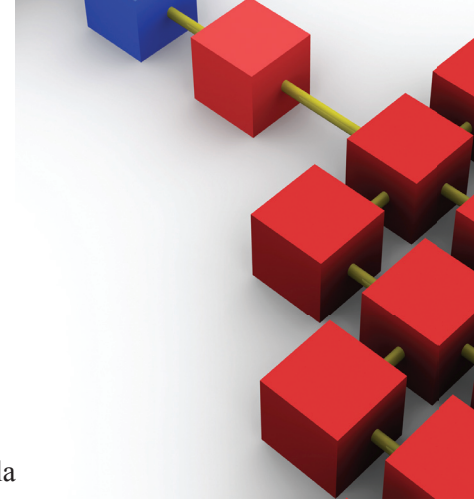
$$= \frac{\left(\frac{1}{4}\right)\left(\frac{2}{5}\right)}{\left(\frac{2}{5}\right)(1) + \left(\frac{1}{4}\right)\left(\frac{2}{5}\right) + \left(\frac{1}{5}\right)\left(\frac{1}{5}\right) + \left(\frac{3}{20}\right)\left(\frac{1}{10}\right)}$$

$$= \frac{\frac{2}{20}}{\frac{2}{5} + \frac{1}{10} + \frac{1}{25} + \frac{3}{200}} = \frac{\frac{1}{10}}{\frac{111}{200}}$$

$$= \frac{200}{1\,110} = 0.18$$

$$= 18\%$$

Que es el mismo resultado que habíamos obtenido.



Capítulo 8

Estadística inferencial

Conceptos básicos

La **probabilidad** razona desde la población hacia la muestra, mientras que la estadística se mueve de la muestra hacia la población.

Por lo regular estamos interesados en obtener razonamientos válidos respecto a datos de un grupo grande de personas o de objetos, pero nuestra capacidad humana para analizar al mismo tiempo grandes cantidades de datos es limitada; además, la mayoría de los problemas estadísticos incluyen mucha información prácticamente imposible de analizar si no se organiza adecuadamente en una tabla de distribución de frecuencias.

La expresión estadística tiene tres connotaciones bien definidas:

A. La estadística, en su acepción común, se considera como una colección de datos numéricos, resultado de observaciones clasificadas y ordenadas según un determinado criterio. Hay estadísticas demográficas, de niveles de producción, de natalidad, de esperanza de vida, de defunciones, entre otras muchas más.

B. La estadística, en una segunda acepción, es el método o la técnica que se sigue para recolectar, organizar, resumir, analizar, generalizar, presentar y contrastar los resultados de fenómenos reales.

C. Por último, la estadística se define como la ciencia que utiliza como instrumento a las matemáticas y al cálculo de probabilidades para estudiar las leyes del comportamiento de fenómenos que dependen del azar (son aleatorios) y no están regidos por leyes físicas.

En una segunda fase, la estadística generaliza leyes y, basándose en ellas, predice o infiere resultados; por ello se le cita como **estadística matemática.**

Conceptos clave

- Probabilidad
- Estadística matemática
- Población
- Métodos estadísticos
- Orden de rango
- Marca de clase
- Diagrama de frecuencia de puntos
- Diagrama de barras
- Histogramas
- Polígono de frecuencias
- Curva de frecuencia
- Ojiva
- Factor de conversión
- Histograma de frecuencia relativa
- Polígono de frecuencia relativa

Población y muestra

En la parte que corresponde al desarrollo del tema de la probabilidad, definimos a la **población** como el conjunto de todos los sucesos susceptibles de aparecer en un problema y que interesan a la persona que hace el estudio. Señalamos también que una muestra es un subconjunto de mediciones seleccionadas de la población.

Ampliaremos estas definiciones para aplicarlas en el estudio de la estadística en la forma siguiente:

La población, según su tamaño, puede ser finita o infinita; y si nos referimos a su número, se toma como el tamaño de la población.

El concepto de *población infinita* sólo existe en teoría porque en la práctica no encontraremos aplicación a poblaciones de elementos infinitos, como sería el caso de las estrellas del universo. Sin embargo, en la estadística matemática las

poblaciones con un número bastante grande de elementos son tratados como si fueran infinitos.

Cuando la población o conjunto es muy grande, se hace difícil la observación de los caracteres a estudiar en cada uno de los elementos, debido al enorme costo que tendría la observación de toda una población y debido también al enorme trabajo y tiempo necesarios para llevar a cabo una observación exhaustiva de cada uno. Estos inconvenientes pueden ser superados mediante la elección de una muestra lo suficientemente representativa de la población.

Problema 1

Necesitamos obtener conclusiones respecto a la situación económica de las familias de 4 500 estudiantes (población) de una escuela de enseñanza media superior, entrevistando únicamente a 50 estudiantes (la muestra) seleccionados al azar.

Problema 2

Queremos obtener conclusiones respecto al porcentaje de focos defectuosos producidos en una fábrica durante una semana de 5 días de trabajo (población), examinando 15 focos diariamente (muestra).

Problema 3

Un partido político desea seleccionar a uno de sus tres miembros más destacados para que participe en las elecciones nacionales. Para ello, decide instaurar un programa de entrevistas en las capitales de tres estados. Se trata de entrevistar a 100 personas (muestra) de manera aleatoria para conocer su preferencia.

Métodos estadísticos

Desde este punto de vista podemos definir la estadística como:

A. *Inferencia estadística o estadística inductiva*. Parte de la ciencia estadística que, con base en los resultados obtenidos del análisis de una muestra de la población, infiere, induce o establece las leyes de comportamiento de la población a la cual pertenece; es también un buen instrumento para aceptar o rechazar las hipótesis que se hayan hecho sobre las características del colectivo de problema.

B. *Estadística descriptiva o deductiva*. Parte de la ciencia estadística que tiene por objeto analizar un determinado conjunto sin pretender obtener conclusiones de tipo más general.

Los métodos y técnicas utilizados por la ciencia estadística tanto en su parte inductiva como en la deductiva son llamados **métodos estadísticos**.

Concepto de variable

En la desigualdad $x + 3 < 6 \;\; x \in z$ los valores enteros que satisfacen esta desigualdad, donde x es la variable, son 2, 1, 0, −1..., que son los valores particulares que puede tomar la x.

Las calificaciones, la altura, el peso, entre otras cosas, son variables relacionadas, por ejemplo, con los alumnos.

Notación

La variable peso, o cualquier otra, se representa en estadística con las letras mayúsculas (... *X*, *Y*, Z). Por ejemplo, cuando nos referimos a *b* estudiantes, cuyos *pesos* en kilogramos son de 43, 45, 60, 50, 52 y 47, decimos que la variable *X* (peso) tiene 6 valores. Para designar estos valores que puede tomar la variable usamos la letra correspondiente en minúscula con un subíndice de orden.

Observa: $(x_1, x_2, \ldots x_n)$; $(y_1, y_2, \ldots y_n)$; $(z_1, z_2, \ldots z_n)$

Variables discretas o continuas

Las variables pueden tener dos valores: uno posible y el otro el realmente observado.

Problema 4

Si en un examen de admisión se aplica a los aspirantes una prueba de opción múltiple con 50 aciertos posibles, el número de aciertos puede tomar 51 valores que van desde 0, 1, 2, 3, 4, ..., 50; éstos son los *valores posibles* que puede tomar la variable *X*, y si los aciertos de 5 alumnos fueron 35, 40, 15, 27, 45, éstos son los v*alores realmente observados de X*.

Una variable es *discreta* cuando entre dos valores reales o posibles no hay ningún otro valor. Es *continua* cuando entre dos valores consecutivos puede haber infinitos valores.

Se entiende por *recorrido* o *rango* (conocido en cálculo como contradominio) de una variable, el espacio de variación numérica de ella; es decir, la diferencia entre el último y el primero de los valores posibles que toma la variable.

Problema 5

Sea *X* el peso en kilogramos de los alumnos de un grupo. Supongamos que el joven más delgado pesa 45 kilos y el más obeso 60 kilos. ¿Cuántos valores posibles de *X* habrá entre 45 y 60 kg? Seguramente serán varios pues habrá quienes pesen 46.450 kilos y otros 48.87 kilos.

El conjunto de números que se asignan al peso de una persona es el conjunto de las fracciones decimales *X*, tales que $45 \leq X \leq 60$.

Podemos decir que no hay ruptura en los valores de *X*; en consecuencia, es una variable continua porque existe una cantidad infinita de valores posibles por cercanos que éstos sean.

Problema 6

Sea *Y* el número de aves de corral en las granjas de una determinada región. Los valores posibles de *Y* son de 0, 1, 2, 3, 4, ..., 350. Una granja puede tener 50 aves, otra 125 y otra más 300. Aceptamos que hay una ruptura, o salto, entre el número de aves de una granja a otra porque la

variable Y (cantidad de aves) no puede tener un valor de, por ejemplo, 125.088 aves; por lo tanto, Y es una variable discreta.

El rango de la variable Y es $0 \leq Y \leq 350 \quad Y \in Z$

Organización de datos

Una vez reunidos los datos de un colectivo para obtener a partir de ellos conclusiones, es necesario organizarlos en una **tabla de distribución de frecuencias**.

La tabla de distribución de frecuencias es una función, ya que cada medida está relacionada con un número que es su frecuencia y, como tal, se puede expresar como una lista, una gráfica o una regla. En estadística se hace con una lista (que es la tabla de frecuencias) o con una gráfica, por ejemplo, un diagrama de frecuencia de puntos.

Las distribuciones de frecuencias de una sola variable se clasifican en tres tipos, según el número de observaciones y el número de valores distintos que toma la variable.

Distribuciones del tipo uno

Son aquellas que constan de un reducido número de observaciones y en consecuencia, de un reducido número de valores distintos a los que toma la variable. Para su presentación no se necesita una técnica determinada, ya que además no es susceptible de tratamiento estadístico porque para que éste exista, es necesario un volumen considerable de observaciones.

Distribuciones del tipo dos

Son las que el número de observaciones es grande, pero el número de valores distintos que toma la variable es pequeño. En este tipo se distribuyen o agrupan los resultados disponibles en dos columnas, una para los valores distintos que toma la variable y otra para la frecuencia de cada uno de ellos.

Problema 7

Para determinar el grado de nutrición de veinte alumnos de secundaria se toma la altura en centímetros de cada uno de ellos:

128	146	136	136	150
140	124	134	142	138
136	120	130	136	132
136	134	142	132	144

Para facilitar su interpretación, los datos se ordenan de forma ascendente o descendente. A este proceso se le llama **orden de rango**.

120	132	136	142
124	134	136	142
128	134	136	144
130	136	138	146
132	136	140	150

Para organizar los datos se usa una tabla de frecuencia que exprese el número de casos de cada categoría (es una distribución del tipo dos).

Altura x_1	Frecuencia n_1	Altura x_1	Frecuencia n_1	Altura x_1	Frecuencia n_1
120	1	130	1	140	1
121	0	131	0	141	0
122	0	132	2	142	2
123	0	133	0	143	0
124	1	134	2	144	1
125	0	135	0	145	0
126	0	136	5	146	1
127	0	137	0	147	0
128	1	138	1	148	0
129	0	139	0	149	2

Distribución del tipo tres

Si el número de observaciones y el número de valores que toma la variable son demasiado grandes para su manejo, las observaciones se agrupan en intervalos $L_{i-1} - L_i$, eligiendo entre ellos una amplitud fija o variable. Los intervalos se anotarán en la primera columna; en la segunda se tabularán los valores para facilitar su conteo y en la tercera se anotará el número de frecuencia (f) correspondiente a cada intervalo.

Los grupos o categorías que incluye $L_{i-1} - L_i$ se llaman *intervalos de clase*; los valores L_{i-1} son los límites inferiores, y L_i los límites superiores de estos intervalos.

Tabla de frecuencias

Clases $L_{i-1} - L_i$	Tabulaciones	Frecuencias (f) n_1
$L_0\ L_1$		n_1
$L_1\ L_2$		n_2
...		
...		
$L_{k-1}\ L_k$		n_k

Problema 8

En un examen departamental de física se examinaron 50 alumnos. Los resultados fueron los siguientes:

87	66	73	68	48
37	76	85	74	65
93	77	66	83	78
49	57	38	69	78
89	96	78	97	74
76	68	63	70	81
64	83	67	61	90
77	88	74	85	80
71	73	61	57	72
80	77	85	80	89

Elabora la tabla de frecuencias.

Solución:

Expresamos los datos en forma ascendente:

37	65	73	78	85
38	66	73	78	85
48	66	74	78	87
49	67	74	80	88
57	68	74	80	89
57	68	76	80	89
61	69	76	81	90
61	70	77	83	93
63	71	77	83	96
64	72	77	85	97

Tabla de frecuencias

Clases $L_{i-1} - L_i$	Tabulaciones	Frecuencias (f) n_1
35 – 39	II	2
40 – 44		0
45 – 49	II	2
50 – 54		0
55 – 59	II	2
60 – 64	IIII	4
65 – 69	~~IIII~~ II	7
70 – 74	~~IIII~~ III	8
75 – 79	~~IIII~~ III	8
80 – 84	~~IIII~~ I	6
85 – 89	~~IIII~~ II	7
90 – 94	II	2
95 – 100	II	2

Al disponer los datos primarios en una distribución del tipo tres, como en la tabla de frecuencias, existe una pérdida de información, ya que no se consideran los resultados obtenidos en forma exacta, sino por aproximación. No se interpretará que un elemento cualquiera tiene un valor cuya medida la da un valor x_i, sino que dicho valor se encuentra entre L_{i-1} y L_i.

Lo que interesa es elegir una amplitud constante o variable lo suficientemente pequeña para que la pérdida sea menor, y al mismo tiempo lo suficientemente grande para que el agrupamiento presente una distribución de no demasiados valores, pues de lo contrario el haber hecho el agrupamiento perderá su finalidad, es decir, la comodidad del manejo. Para facilitar el cálculo es recomendable escoger estos intervalos de manera que sus puntos medios sean múltiplos de números como el 5 o como el 10, y generalmente no debe haber menos de 7 intervalos ni más de 15, aunque no hay normas fijas.

No es necesario que los intervalos de clase sean iguales. Cada persona elige el intervalo de clase más adecuado de acuerdo con lo que desea investigar.

Marca de clase

Una vez hecho todo lo anterior, y antes de aplicar los métodos estadísticos, es necesario sustituir cada intervalo por un número. A este número se le llama **marca de clase** y es el valor central de cada intervalo; es decir, la media aritmética de los límites inferior y superior. La marca de clase se obtiene así:

marca de clase: $x_i = \dfrac{L_{i-1} + L_i}{2}$ se abrevia (m.c.)

Tabla de frecuencias

Clases $L_{i-1} - L_i$	Tabulaciones	Frecuencias (f) n_i	Marca de clase (m.c.) x_i
$L_0 - L_1$		n_1	x_1
$L_1 - L_2$		n_2	x_2
...			
...			
$L_{n-1} - L_n$		n_n	x_n

Problema 9

Determina las marcas de clase del problema anterior.

$\dfrac{35+39}{2} = \dfrac{74}{2} = 37$

$\dfrac{40+44}{2} = \dfrac{84}{2} = 42$

$\dfrac{45+49}{2} = \dfrac{94}{2} = 47$

$\dfrac{70+74}{2} = \dfrac{144}{2} = 72$

$\dfrac{75+79}{2} = \dfrac{154}{2} = 77$

$\dfrac{80+84}{2} = \dfrac{164}{2} = 82$

$$\frac{50+54}{2}=\frac{104}{2}=52$$

$$\frac{55+59}{2}=\frac{114}{2}=57$$

$$\frac{60+64}{2}=\frac{124}{2}=62$$

$$\frac{65+69}{2}=\frac{134}{2}=67$$

$$\frac{85+89}{2}=\frac{174}{2}=87$$

$$\frac{90+94}{2}=\frac{184}{2}=92$$

$$\frac{95+100}{2}=\frac{195}{2}=97.5$$

Los datos ya obtenidos los anotamos en la tabla de frecuencias del problema anterior de la siguiente manera:

Tabla de frecuencias

Clases $L_{i-1} - L_i$	Tabulaciones	Frecuencias (f) n_1	Marca de clase (m.c.)
35 – 39	II	2	37
40 – 44	0	0	42
45 – 49	II	2	47
50 – 54	0	0	52
55 – 59	II	2	57
60 – 64	IIII	4	62
65 – 69	~~IIII~~ II	7	67
70 – 74	~~IIII~~ III	8	72
75 – 79	~~IIII~~ III	8	77
80 – 84	~~IIII~~ I	6	82
85 – 89	~~IIII~~ II	7	87
90 – 94	II	2	92
95 – 100	II	2	97.5

Al poner la marca de clase se comete un *error de agrupamiento* porque las n_i (frecuencias "f") no son las veces que se repite el valor de x de la variable, sino que son las veces que aparecen valores de la variable considerados entre L_{i-1} L_i.

Aceptamos que la pérdida de información a que nos referimos y este error de agrupamiento son absolutamente necesarios para que las distribuciones del tipo tres puedan recibir un tratamiento estadístico.

Gráficas

La adecuada representación gráfica de una distribución de datos ayuda a obtener conclusiones sobre el comportamiento real de la variable. En este caso, es necesario que el impacto visual de la representación corresponda a la realidad y en consecuencia,

que el método aplicado se base en principios geométricos válidos. Sin embargo, no debemos olvidar que una gráfica no pasa de ser un instrumento auxiliar de análisis.

Muchas formas de representación gráfica que aparecen en la sección económica de los diarios informativos son presentaciones de distribuciones del tipo uno, tales como las variaciones de los precios del petróleo en un mes, la representación del precio del dólar en una semana, etcétera.

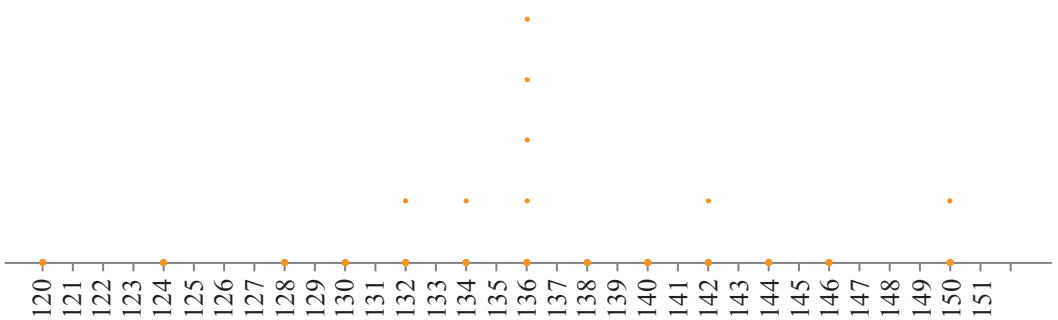
A continuación analizaremos el diagrama de frecuencia de puntos, el diagrama de barras, el histograma y el polígono de frecuencias, los cuales se usan para presentaciones de distribuciones del tipo dos y tres.

Diagrama de frecuencia de puntos

Es la información gráfica de cómo están distribuidos los datos sobre el rango (contradominio en cálculo).

Problema 10

Representa gráficamente la solución del problema 7 con un diagrama de frecuencia de puntos.

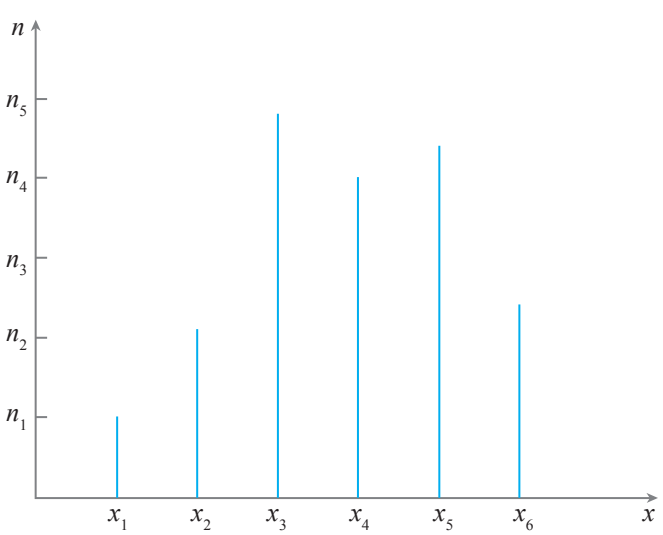


Observa que las alturas se agrupan alrededor de 136.

Diagrama de barras

Se usa cuando se dispone de muchas observaciones pero pocos valores de la variable (distribución del tipo dos).

Este diagrama se elabora señalando en el eje de las x (abscisas) de un sistema de ejes coordenados, los valores de la variable, colocando sobre ellas unas columnas a escala de las alturas igual a la frecuencia de cada uno de los valores medidos en el sentido del eje de las y (ordenadas).

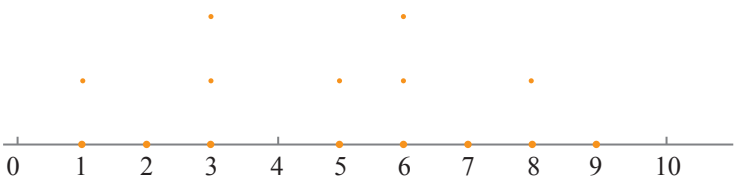


Problema 11

Un grupo de 15 alumnos presenta examen extraordinario de química; un funcionario de la escuela necesita saber cuántos alumnos obtuvieron calificación inferior a 6 y cuántos entre 6 y 8.

X	f
0 puntos	0
1 punto	2
2 puntos	1
3 puntos	3
4 puntos	0
5 puntos	2
6 puntos	3
7 puntos	1
8 puntos	2
9 puntos	1
10 puntos	0

Para resolver este tipo de problemas, ordenamos las calificaciones en una tabla de frecuencias y contestamos preguntas como "inferior o igual que" y "superior a". Así:



De donde 8 alumnos obtuvieron una calificación menor a 6 y su calificación está entre 6 y 8.

Histogramas. Datos agrupados

El histograma es la gráfica más común y se utiliza cuando el número de observaciones y el número de valores que toma la variable son grandes (distribución del tipo tres).

Los histogramas son una forma de representar las frecuencias de clase por medio de áreas rectangulares (barras), pero son diferentes a los diagramas de barras, cuyas alturas miden el tamaño de la variable y generalmente se dibujan separadas, dejando espacios entre ellas. En los histogramas, las frecuencias quedan representadas por el área de los rectángulos, no por sus alturas; y las barras se dibujan sin dejar espacios entre ellas.

Concepto de densidad

La densidad física es un concepto relativo que relaciona la masa de un cuerpo con su volumen. En estadística, por la densidad de frecuencia, se obtiene la frecuencia absoluta o número de casos que hay dentro del intervalo de clase.

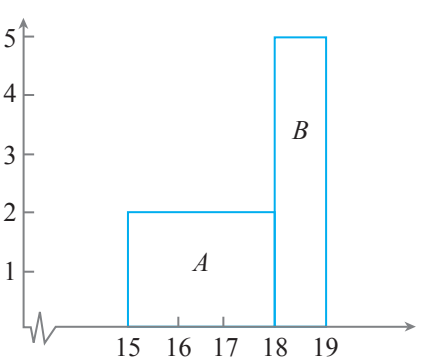
En los histogramas, el eje vertical mide la densidad de frecuencias y el eje horizontal mide los intervalos de clase. Así:

El histograma de la figura corresponde a intervalos de clase de diferente anchura, el rectángulo A representa el intervalo 15-18; en el B, el rectángulo es 18-19.

Área rectángulo $A = 3\ (2) = 6$

Área rectángulo $B = 1\ (5) = 5$

El rectángulo A representa 6 unidades de frecuencia y el B 5 unidades. Observa que la base del rectángulo B es de una unidad (no importa de qué ancho la hayas medido) y la altura de B es de 5 unidades; la altura corresponde a la frecuencia.



Considera siempre la frecuencia como una expresión del área de los rectángulos.

Longitud de los ejes para expresar un histograma

Para la elección de la longitud de los ejes te recomendamos aplicar la regla de los tres cuartos, la cual señala:

El eje vertical debe ser tres cuartos de la longitud del eje horizontal.

El eje horizontal de las abscisas se elige de acuerdo con la necesidad del problema.

La misma unidad que usaste para dividir el eje horizontal la usarás para dividir el eje vertical.

Problema 12

Traza el histograma de la distribución de frecuencias agrupadas siguientes:

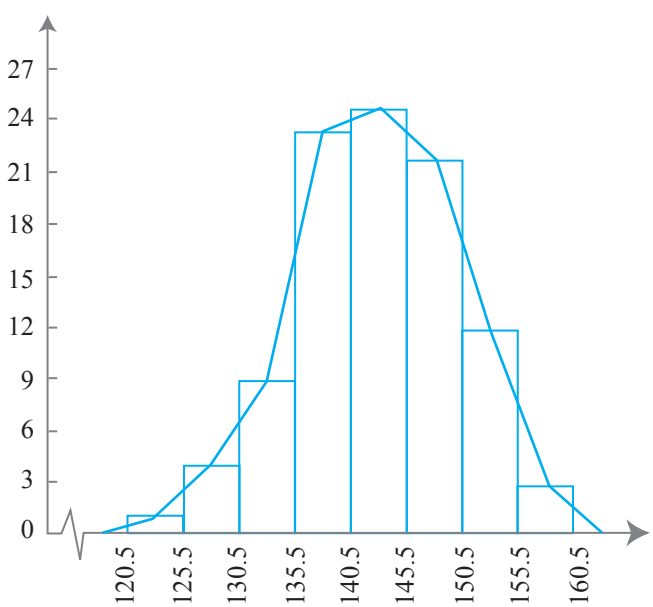
Clases	Frecuencia
120.5 – 125.5	1
125.5 – 130.5	4
130.5 – 135.5	9
135.5 – 140.5	23
140.5 – 145.5	25
145.5 – 150.5	22
150.5 – 155.5	12
155.5 – 160.5	3

Para trazar el histograma procedemos así:

1. Sobre el eje de las abscisas ponemos a escala los valores de la variable X (los puntajes), en nuestro ejemplo será de 7 milímetros.
2. Se trazan perpendiculares sobre el eje horizontal de la longitud que sea necesaria; por ejemplo, a partir del punto que corresponde al 125.5 límite superior de la primera clase se traza una línea cuya longitud sea una unidad, frecuencia correspondiente a la clase, en 130.5 una de 4, en 135.5 una de 9, ...,

Observa que al trazar la perpendicular en 135.5 se prolonga y así para todas después de la primera.

Solución:



Polígonos de frecuencias

El polígono de frecuencias se obtiene uniendo los puntos medios de los intervalos de clase del histograma, como lo hicimos en la figura anterior.

Aun cuando no hay frecuencia antes del primer y después del último intervalo de clase se continúa el polígono.

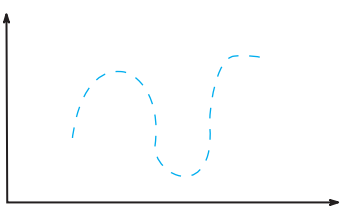
Se puede demostrar trigonométricamente que el área del polígono de frecuencias es igual a la del histograma.

Curvas de frecuencia

Observamos cómo de los histogramas se evoluciona hacia los contornos poligonales. Suavizando los polígonos se obtienen curvas que, en estadística, son de utilidad para visualizar propiedades y analizar, comparar y obtener información.

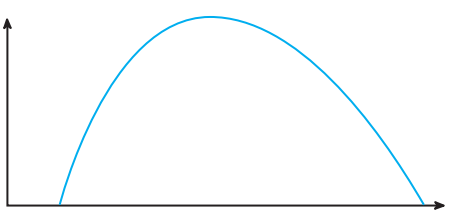
Los datos que se procesan en estadística pertenecen a muestras aleatorias de una población o universo que incluye un gran número de elementos. Si las variables son continuas, teóricamente, para cualquier intervalo por pequeño que éste sea habrá algunas observaciones que estén dentro de él.

Si pensamos que el polígono de frecuencias está formado por pequeños segmentos se aproximará a una curva, por ello se da el nombre de curva a cualquier polígono.

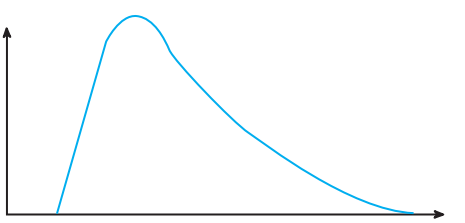


Se acostumbra trazar a mano conservando el patrón del histograma.

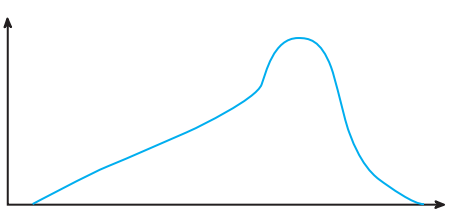
Si se usan curvas en lugar de histogramas, se pueden expresar en un mismo gráfico sin que se confundan. Esta tarea no se puede realizar con los histogramas.



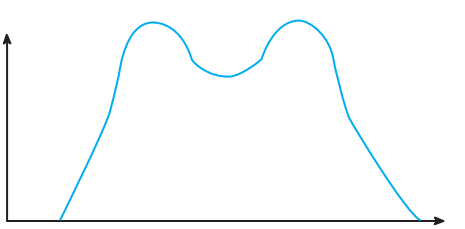
Simétrica, curva normal



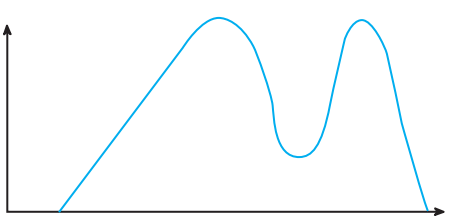
Asimétrica, sesgada a la izquierda



Asimétrica, sesgada a la derecha



Simétrica bimodal



Asimétrica bimodal

Problema 13

Representa gráficamente con curvas de frecuencia las siguientes distribuciones de frecuencias agrupadas.

a)

Clases	Frec.
22.5 – 25.5	2
25.5 – 28.5	8
28.5 – 31.5	9
31.5 – 34.5	8
34.5 – 37.5	6
37.5 – 40.5	2
40.5 – 43.5	1

b)

Clases	Frec.
23.5 – 28.5	3
28.5 – 33.5	7
33.5 – 38.5	3
38.5 – 43.5	8
43.5 – 48.5	2

c)

Clases	Frec.
24.5 – 27.5	2
27.5 – 30.5	5
30.5 – 33.5	6
33.5 – 36.5	5
36.5 – 39.5	2

Solución:

Para localizar los puntos y trazar las curvas de frecuencia es indispensable calcular las marcas de clase como ya se hizo para el histograma.

Para 22.5 – 25.5

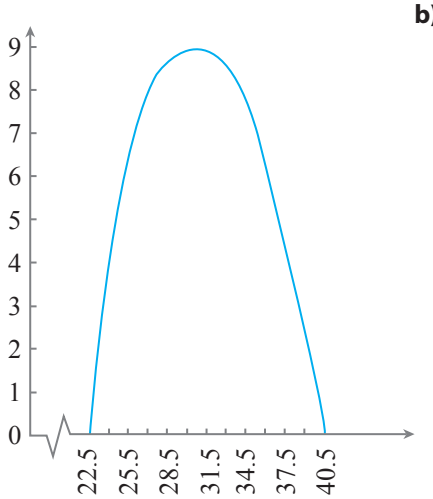
$$\frac{22.5+25.5}{2}=\frac{48}{2}=24$$

$$\frac{25.5+28.5}{2}=\frac{54}{2}=27$$

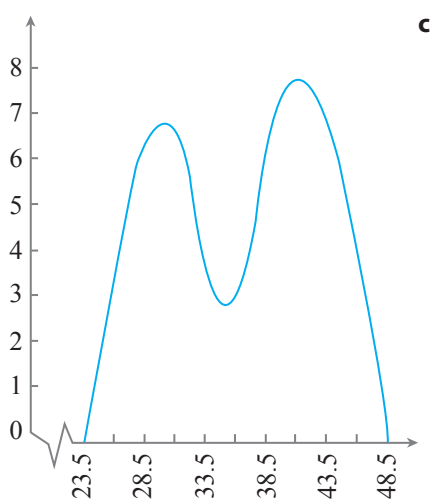
y así sucesivamente para todas las demás que calcularemos sin expresar las operaciones.

Trazamos las gráficas.

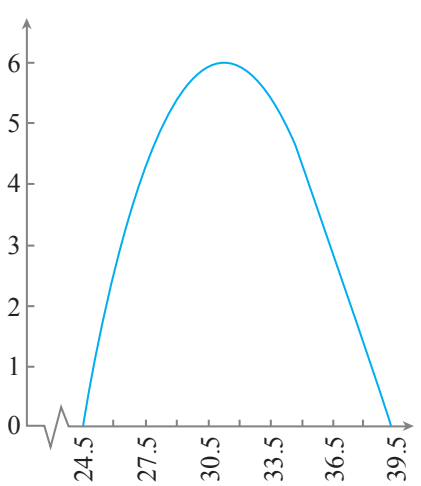
a)



b)



c)



Cuadro de distribución de frecuencias agrupadas:

Clases	Frecuencias
123.5 – 128.5	4
128.5 – 133.5	4
133.5 – 138.5	8
138.5 – 143.5	21
143.5 – 148.5	6
148.5 – 153.5	25
153.5 – 158.5	21
158.5 – 163.5	10
163.5 – 168.5	1
Total	$N - 100$

Frecuencias acumuladas. Ojivas

El cuadro a la izquierda expresa la distribución de frecuencias agrupadas no acumulativas que se elaboró en una escuela, donde se tomó para fines de control nutricional la estatura en centímetros de 100 alumnos.

Algunas veces es necesario, conocer cuántas observaciones caen por debajo de cierta puntuación o por encima de ella. Esta información se obtiene con los cuadros de frecuencias acumuladas.

Problema 14

Con base en el cuadro anterior de distribución de frecuencias agrupadas, determina dos cuadros: uno de frecuencias acumuladas hacia abajo y otro de frecuencias acumuladas hacia arriba, y traza las ojivas correspondientes.

Solución:

Cuadro A

Estatura	Número de alumnos
123.5	0
128.5	4
133.5	8
138.5	16
143.5	37
148.5	43
153.5	68
158.5	89
163.5	99
168.5	100

Frecuencia acumulada de estaturas que expresa el número de alumnos que miden menos de la estatura indicada.

Para completar el cuadro de distribución acumulada observamos en el cuadro de distribución de frecuencias que para menos de 123.5 hay 0; para menos de 128.5 son 4; de 133.5 hay 8, etcétera.

Cuadro B

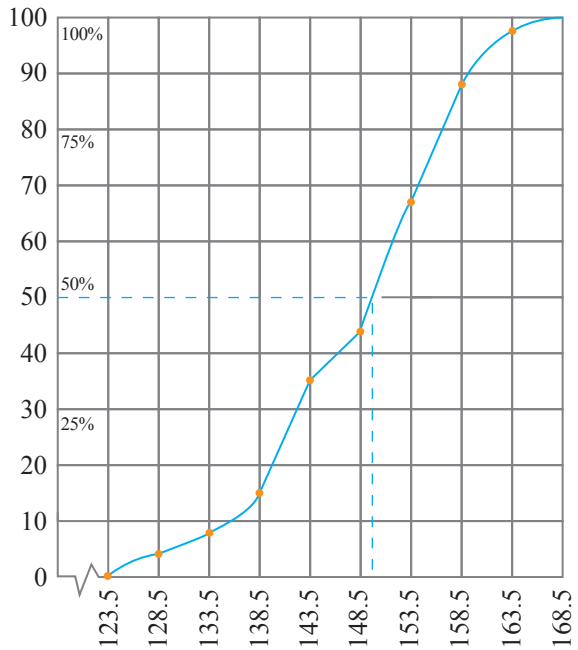
Estatura	Número de alumnos
123.5	100
128.5	96
133.5	92
138.5	84
143.5	63
148.5	57
153.5	32
158.5	11
163.5	1
168.5	0

Frecuencia acumulada de estaturas que expresa el número de alumnos que miden más de la estatura indicada.

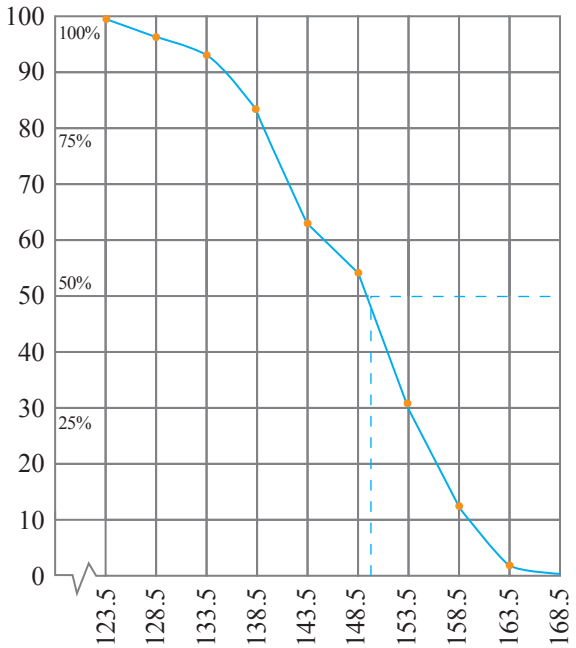
Para completar el cuadro de distribución acumulada observamos en el cuadro de distribución de frecuencias que por encima de 123.5 hay 100; de 128.5 son 96; de 133.5 es de 92, etcétera.

La representación gráfica de la información que incluye los cuadrados de frecuencias acumuladas se hace con las curvas llamadas **ojivas**.

Una ojiva incluye un eje horizontal donde se expresan las alturas sobre los intervalos y un eje vertical donde se expresan las frecuencias acumuladas y los porcentajes acumulados.



Ojiva *A* (menos que)



Ojiva *B* (más que)

Las líneas puntedas representan la *mediana*.

Distribuciones de frecuencias relativas

Organizar la información en una tabla de frecuencias, presentarla en cuadros, marcar los intervalos de clase y hacer las gráficas de frecuencias absolutas permiten relacionar y comprender los valores de un mismo colectivo.

A continuación relacionaremos observaciones sobre un mismo aspecto de dos colectivos diferentes, lo que sólo es posible si las observaciones tienen la misma base.

Consideremos el caso siguiente:

Juan obtuvo en su examen extraordinario de matemáticas 30 puntos de una calificación sobre base 70 y en física obtuvo 35 puntos de una calificación de base 50.

Las calificaciones de 30 y 35 se dan en puntuaciones absolutas y por eso no pueden compararse, ya que intrínsecamente como valores no tienen significado: su verdadero sentido surge cuando se da el valor relativo a las bases de calificaciones: en matemáticas sobre base 70 y en física sobre base de 50.

Con las calificaciones y las bases se pueden establecer proporciones y lograr comparaciones en valores relativos y no en valores absolutos como se citaron; si elegimos la base 100 para las dos, obtendremos las calificaciones relativas en tanto por ciento:

Para matemáticas	Para física
$30 : 70 :: x : 100$	$35 : 50 :: x : 100$
$70x = 30(100)$	$50x = 35(100)$
$x = \frac{3\,000}{70}$	$x = \frac{3\,500}{50}$
$x = 42.86\%$	$x = 70\%$

Ahora tenemos una buena información y mejor si se da la posición relativa de las calificaciones de Juan, con relación al grupo de alumnos en que se examinó, en la forma siguiente:

En el examen de matemáticas de un total de 40 alumnos, 16 de ellos obtuvieron calificación igual o menor que 30, Juan estaría en el nivel $\frac{16}{40} = 0.40$; si este resultado lo multiplicamos por 100 (nueva base) se obtiene 40% y Juan pensará: el 60% obtuvo mejor calificación en matemáticas que yo.

Y en física, de un total de 25 alumnos, 15 lograron calificación igual o menor que 35 puntos, Juan estará en el nivel $\frac{15}{25} = 0.60$. Si este resultado lo multiplicamos por la nueva base, se obtiene 60% y Juan pensará: el 40% obtuvo mejor calificación en física que yo.

Juan concluyó: es probable que apruebe en física y repruebe en matemáticas.

La frecuencia relativa de una clase se obtiene en tanto por ciento, que es la nueva base, si dividimos la frecuencia de la clase entre el número total de frecuencias y el resultado lo multiplicamos por 100 (como ya se hizo en el caso de Juan).

$$\text{Frecuencia relativa} = \frac{L}{N}(100)$$

Con base en los datos de un cuadro de distribución de frecuencias agrupadas, se incluyen en él las correspondientes frecuencias relativas, como se acostumbra en la práctica incluir en un mismo cuadro las frecuencias absolutas y las relativas.

Para facilitar el cálculo de las frecuencias relativas de cada clase, se usa un **factor de conversión**, que resulta de dividir 100 por el número total de frecuencias.

$$\text{Factor} = \frac{100}{N}$$

Problema 15

Las autoridades de la Secretaría de Educación Pública deciden que en otra escuela también se tomen las estaturas en centímetros de todos los alumnos, pero ahora de los menores de 17 años.

Elabora un cuadro de frecuencias agrupadas que incluya las frecuencias absolutas y las relativas, estas últimas en tanto por ciento.

Cuadro de distribución de frecuencias agrupadas

Clase	Frecuencias	
	Absolutas	Relativas
123.5 – 128.5	2	1.639
128.5 – 133.5	3	2.459
133.5 – 138.5	8	6.557
138.5 – 143.5	20	16.393
143.5 – 148.5	9	7.377
148.5 – 153.5	8	6.557
153.5 – 158.5	30	24.590
158.5 – 163.5	23	18.852
163.5 – 168.5	15	12.295
168.5 – 173.5	4	3.279
Totales	122	100.00

Nota: La suma exacta de las frecuencias relativas de las cantidades citadas en el cuadro es de 99.998. Las 0.002 milésimas que faltan para 100 se originaron porque en la división de $\frac{100}{122}$ únicamente tomamos 3 cifras decimales (redondeadas).

Para calcular las frecuencias relativas previamente obtuvimos el *factor de conversión*.

Factor = $\frac{100}{122}$ = 0.820 (se tomaron 3 cifras redondeadas)

2(0.820) = 1.639

3 (0.820) = 2.459

y así con las demás.

Los histogramas y los polígonos de frecuencias se convierten en **histogramas de frecuencia relativa** y **polígonos de frecuencia relativa** al cambiar la escala vertical de valores absolutos por otra de valores relativos o proporcionalidades de base 100 (escala en tanto por ciento).

Distribuciones porcentuales acumuladas

Los cuadros de frecuencia acumulada porcentuales se obtienen de manera semejante que los cuadros de frecuencia acumulada, es decir, convirtiendo las frecuencias

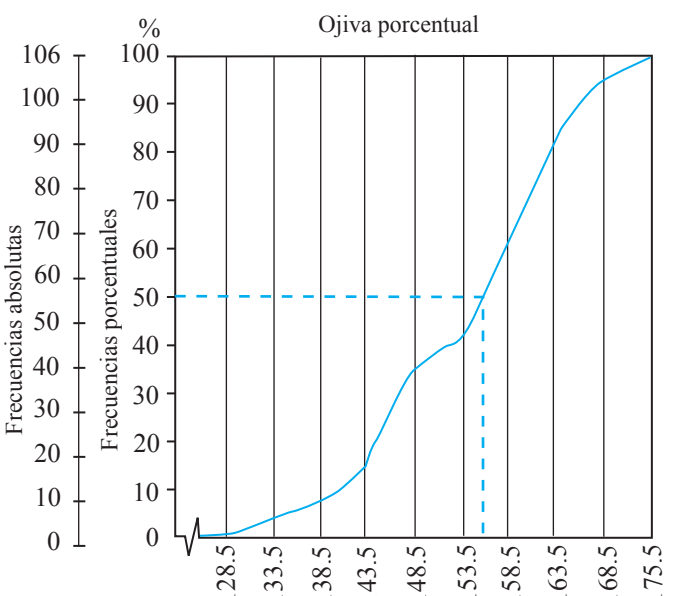
acumuladas en frecuencias relativas o proporcionalidades de base 100. Se acostumbra citar las dos columnas en un mismo cuadro.

Problema 16

En el siguiente cuadro de distribución acumulativa de estaturas de un grupo de alumnos, que expresa el número de ellos que están por debajo de la estatura indicada, agrega la columna que corresponde a las frecuencias relativas y traza la ojiva porcentual.

Solución:

Estatura	Frecuencias	
	Absolutas	Relativas
128.5	0	0.000
133.5	2	1.886
138.5	5	4.715
143.5	14	13.202
148.5	38	35.834
153.5	45	42.435
158.5	65	61.295
163.5	89	83.927
168.5	103	97.129
173.5	106	100.000



Para completar el cuadro, primero calculamos el factor de conversión:

$$\text{Factor} = \frac{100}{106} = 0.943$$

y obtuvimos las frecuencias relativas:

0(0.943) = 0.000

2(0.943) = 1.886

5(0.943) = 4.715

y así con las demás.

El eje porcentual se divide de 10 en 10. La línea punteada representa la *mediana*.

Percentiles y rango percentil

La expresión *percentil* se usa para indicar en una distribución de observaciones el valor por debajo del cual está situado cierto porcentaje de distribución de valores; por ejemplo, al decir que en una distribución de estaturas el 15.28% de los alumnos mide 144.5 o menos, se expresa:

$P_{15.28} = 144.5$ entonces estamos afirmando que el 15.28% de los alumnos está por debajo de 144.5 centímetros de estatura.

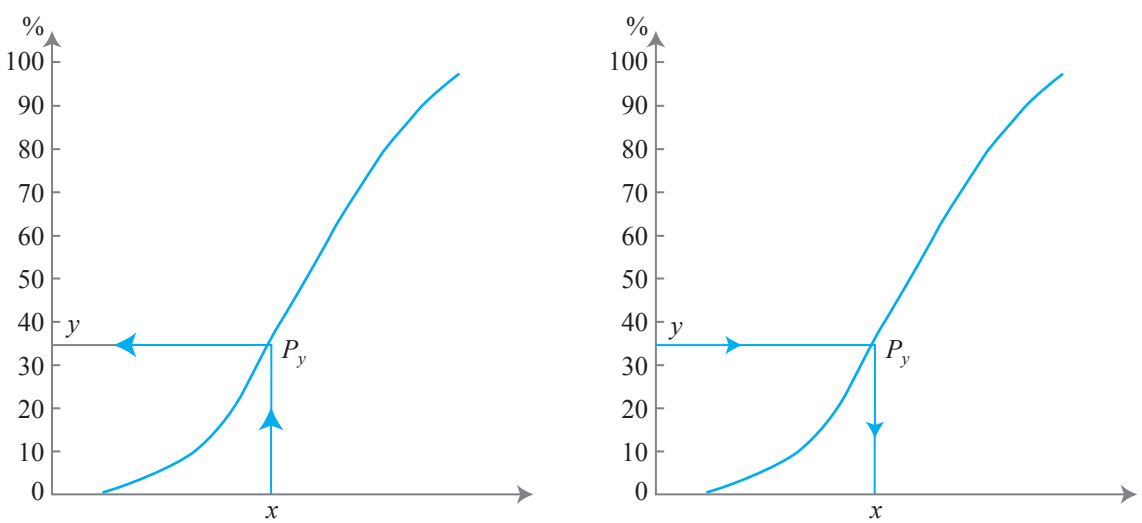
Se presentan dos problemas relacionados al uso de los percentiles:

a) Obtener el valor de la abscisa x que corresponde a un valor percentil, y

b) Obtener el rango percentil correspondiente a un valor de la abscisa x.

Solución utilizando la gráfica de la ojiva:

- **1.** Si conocemos el valor de x obtenemos el rango percentil y procedemos así: En la gráfica de la ojiva se traza, por el punto x conocido, una paralela al eje de las ordenadas hasta intersecar la ojiva, y desde el punto de intercepción se traza una paralela al eje de las abscisas y obtenemos el rango percentil P_y.
- **2.** Si conocemos el percentil (valor de y), determina el valor de la abscisa x. Se traza por el punto que corresponde al percentil y (P_y) una paralela al eje de las abscisas hasta intersecar la ojiva; desde el punto de intercepción se baja una perpendicular al eje de las x .



Ejercicios de repaso

En los dos problemas siguientes señala cuál es la población y cuál es la muestra en cada uno de ellos.

1. Un inspector de zona escolar necesita conocer si los alumnos de una escuela primaria con 760 niños reciben el contenido de los programas oficiales de aritmética y decide entrevistar a 85 alumnos seleccionados al azar.

2. Para conocer la preferencia política de los habitantes de una población de 750 000 habitantes mayores de 18 años, un funcionario de la Secretaría de Gobernación decide que se entreviste al azar un número no menor de 8 000 personas.

3. En un examen departamental de química se examinaron 60 alumnos con los resultados siguientes:

90	63	64	68	56	38
43	69	37	60	80	29
75	82	83	45	67	67
75	30	43	82	43	45
67	68	70	75	42	55
89	74	76	81	83	85
88	85	90	89	43	32
32	34	80	79	83	84
45	65	77	78	90	89
39	35	43	44	65	67

a) Expresar los datos en forma ascendente.

b) Elabora con ellos una tabla de frecuencias.

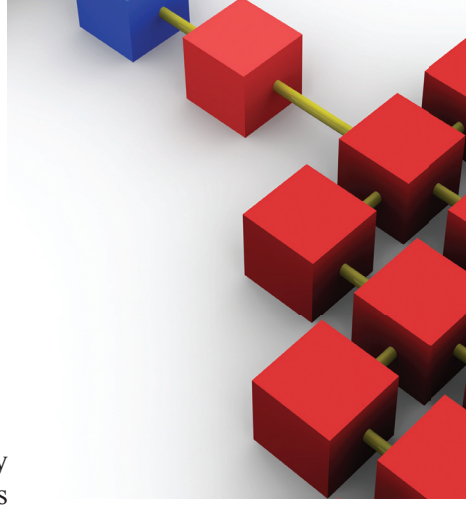
4. Representa gráficamente con curvas de frecuencia las siguientes distribuciones de frecuencias agrupadas.

a)

Clases	Frecuencias
22.5 – 27.5	3
27.5 – 32.5	5
32.5 – 37.5	9
37.5 – 42.5	7
42.5 – 47.5	2
47.5 – 52.5	3
52.5 – 57.5	1

b)

Clases	Frecuencias
39.5 – 42.5	2
42.5 – 45.5	8
45.5 – 48.5	5
48.5 – 51.5	2
51.5 – 54.5	9
54.5 – 57.5	3



Capítulo 9

Medidas de tendencia central

Generalidades

En el capítulo anterior analizamos la información estadística con histogramas y polígonos de frecuencias. Observamos un significativo comportamiento de los datos, en cuanto a la frecuencia en que se presentan los valores, ya que algunos de éstos son más comunes que otros. Notamos una clara tendencia de agrupación en un entorno de los valores más frecuentes, de manera que las curvas representativas toman forma de campana; y por lo general, la mayor densidad de frecuencia está en la parte central de las gráficas, de ahí el nombre de medidas de tendencia central que se da a la media aritmética, la mediana y la moda. Además, nos referiremos a la media geométrica y a la media armónica.

Conceptos clave

- Parámetro
- Valores o resultados estadísticos
- Media aritmética
- Método de mínimos cuadrados
- Mediana
- Moda
- Moda de datos agrupados
- Media geométrica
- Media armónica

Parámetro

En estadística se denomina **parámetro** a cualquier característica cuantificable de una población y se les designa con letras griegas. A los resultados obtenidos de operar los datos resultantes de una muestra se llaman **valores** o **resultados estadísticos**; y a partir de éstos se obtienen los parámetros de la población, mismos que se designan con letras minúsculas de nuestro alfabeto.

Media aritmética

Es la suma de los valores de cierto número de cantidades dividido entre su número.

Se expresa así:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{N}$$

Donde:

N: Es el número de observaciones

X: El valor de cada observación

$\overline{X}$: Es la media aritmética, media, o X barra

La media es la única de las medidas de tendencia central que puede intervenir en operaciones algebraicas.

De sus propiedades nos referiremos a las dos siguientes:

Primera

Si X es una de las variables, su desviación respecto a $\overline{X}$ es la diferencia $X - \overline{X}$. La suma de estas diferencias es cero.

$$\sum_{i=1}^{n} \left(X_i - \overline{X} \right) = 0$$

En toda distribución, la suma de las desviaciones de cada uno de los valores de la variable respecto a la media es cero.

Segunda

La suma de los cuadrados de las desviaciones respecto a la media es siempre menor que la suma de los cuadrados de las desviaciones con respecto a cualquier otro valor.

Propiedad que indica que la media es la medida de tendencia central que hace mínima la suma de los cuadrados de las desviaciones en un entorno de donde:

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)=0$$

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)=\sum_{i=1}^{n}X-\sum_{i=1}^{n}\overline{X}$$

Y como:

$$\sum_{i=1}^{n}\overline{X}=\underbrace{\overline{X}+\overline{X}+\overline{X}+\ldots+\overline{X}}_{n\text{ veces}}$$

Entonces:

$$N\overline{X}_i=\sum_{i=1}^{n}X_1 \text{ por definición y}$$

$$\sum_{i=1}^{n}\overline{X}=N\overline{X}$$

Se tiene:

$$\sum_{i=1}^{n}X-\sum_{i=1}^{n}\overline{X}=N\overline{X}-N\overline{X}=0$$

Esta propiedad da origen al llamado **método de mínimos cuadrados** para calcular la media y es importante en estadística por su aplicación en el ajuste de curvas.

Problema 1

Determina la media del precio del petróleo que se registró en un mes si se vendió en el mercado mundial en 28, 31, 29, 27, 26 dólares por barril.

Solución:

$$\overline{X}=\frac{28+31+29+27+26}{5}=\frac{141}{5}$$

$$\overline{X}=28.2$$

Problema 2

En un examen extraordinario las calificaciones obtenidas por un grupo de 13 alumnos sobre un máximo de 10 puntos, fueron:

En matemáticas: 4, 7, 3, 6, 2, 8, 4, 7, 0, 1, 7, 6, 4

En física: 8, 4, 3, 6, 7, 5, 6, 2, 1, 7, 6, 7, 0

Calcula el promedio de matemáticas, de física y el de ambas materias.

Solución:

$$\overline{X}=\frac{4+7+3+6+2+8+4+7+0+1+7+6+4}{13}=\frac{59}{13}=4.54 \text{ matemáticas}$$

$$\bar{X} = \frac{8+4+3+6+7+5+6+2+1+7+6+7+0}{13} = \frac{62}{13} = 4.77 \text{ física}$$

Aceptamos que la media aritmética reparte entre todos la suma total de los valores observados, dando a cada uno el mismo valor promedio; en consecuencia, si conocemos la media y el número N de observaciones para calcular la suma de todos los valores, multiplicamos la media por N:

$$\sum_{i=1}^{n} \bar{X}_i = N\bar{X} = 13(4.54) = 59.02$$

La diferencia de 59 y 59.02 resulta porque al dividir $\frac{59}{13} = 4.54$ los centésimos se redondearon a 4.

Problema 3

En una pequeña fábrica maquiladora el promedio del sueldo de ocho trabajadores es de $875.00 a la quincena.

Calcula el monto de la nómina mensual.

Solución:

Valor de la nómina mensual = 2(8) (875) = $14 000.00

Media aritmética de una distribución de frecuencias agrupadas

La media aritmética ponderada se aplica para calcular el valor promedio de cantidades a cada una de las cuales está asociado un número que la pondera.

Se expresa:

Si X_1, X_2, …X_n son las cantidades y m_1, m_2, …m_n las respectivas ponderaciones, entonces la media ponderada $\bar{X}$ es:

$$\bar{X} = \frac{m_1X_1m_2X_2 + \ldots m_3X_n}{m_1 + m_2 + \ldots + m_n} = \frac{\sum_{i=1}^{n} m_iX_i}{\sum_{i=1}^{n} m_i}$$

Problema 4

Si un comerciante compra dos partidas de camisas: una de 60 a $75.00 cada una y otra de 30 en $83.50 cada una, determina el precio promedio de cada camisa.

Solución:

$$\text{Precio promedio} = \frac{60(75)+30(83.50)}{60+30} = \frac{4\,500 + 2\,505}{90} = \frac{7\,005}{90}$$

$$= \$77.83 \text{ (se tomaron dos cifras decimales)}$$

Problema 5

La tendencia actual para ingresar a una licenciatura exige que el aspirante tenga un promedio de 8.5 y apruebe el examen de admisión con 6 como

mínimo. No todas las materias que se evalúan en el examen tienen el mismo valor, es decir, cada una tiene una ponderación diferente.

Un aspirante obtuvo las calificaciones siguientes: matemáticas 8, física 7, español 4, inglés 6; para averiguar si el alumno ingresa a la universidad se tiene que calcular el promedio ponderado. Las ponderaciones son: matemáticas 7, física 7, español 3, inglés 5.

Solución:

$$\bar{X} = \frac{8(7)+7(7)+4(3)+6(5)}{7+7+3+5} = \frac{56+49+12+30}{22} = \frac{147}{22}$$

$$= 6.68$$

Para calcular la media aritmética de una distribución de frecuencias agrupadas tomamos en cuenta que a todos los valores que hay dentro de un intervalo de clase se les considera de un mismo valor igual a la marca de clase, y las frecuencias son las ponderaciones de los valores en correspondencia con las marcas de clase y la suma de las frecuencias es el total de veces que se tiene registro.

Se expresa:

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}$$

Problema 6

Calcula la media aritmética de la distribución de frecuencias agrupadas de la tabla de frecuencias que obtuvimos al resolver el problema 8 del capítulo anterior y que citamos a continuación.

Clases $L_{i=1} - L_i$	Tabulaciones n_i	Frecuencias (f)	Marca de clase $(m, c)\ x_i$
35 – 39	II	2	37
40 – 44	0	0	42
45 – 49	II	2	47
50 – 54	0	0	52
55 – 59	II	2	57
60 – 64	IIII	4	62
65 – 69	𝍸 II	7	67
70 – 74	𝍸 III	8	77
75 – 79	𝍸 III	8	77
80 – 84	𝍸 I	6	82
85 – 89	𝍸 II	7	87
90 – 94	II	2	92
95 – 100	II	2	97.5

Solución:

Para obtener la media aritmética procedemos así:

Intervalos	Marca X	Frecuencia (f)	$f(x)$
35 – 39	37	2	74
40 – 44	42	0	0
45 – 49	47	2	94
50 – 54	52	0	0
55 – 59	57	2	114
60 – 64	62	4	248
65 – 69	67	7	469
70 – 74	72	8	576
75 – 79	77	8	616
80 – 84	82	6	492
85 – 89	87	7	609
90 – 94	92	2	184
95 – 100	97.5	2	195

$$\sum_{i=1}^{n} f_i = 50 \qquad \sum_{i=1}^{n} f_i X_i = 3\,671$$

Sustituimos en:

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i} = \frac{3\,671}{50} = 73.4$$

El valor de la media obtenida de la frecuencia agrupada es suficientemente aproximado para trabajos de estadística, ya que al trabajar con datos agrupados se pierde parte de la información primaria y no hay otro recurso más que trabajar con marcas de clases en lugar de los datos originales. Además, el valor de la media no será suficientemente aproximado si la distribución de frecuencias agrupadas es muy irregular o demasiado asimétrica.

Mediana y moda

La mediana y la moda son medidas de tendencia central que por sus propiedades destacan los valores individuales de un colectivo; en cambio, la media aritmética, al promediar todos los valores igualando en un justo reparto todas las observaciones, suprime sus individualidades.

Mediana

Se define como el valor que divide un conjunto de datos previamente ordenados de menor a mayor. Es además el punto intermedio entre todos ellos.

Si el número N de datos es impar, entonces hay un número intermedio; por ejemplo, si tenemos cinco datos: 3, 5, 7, 9, 11, el número 7 es el punto intermedio. Si el

número N de datos es par, entonces hay dos datos intermedios; por ejemplo, en la media de los valores 8, 10, 16, 19, 23, 25, hay dos valores centrales, que son 16 y 19; el valor equidistante entre ellos es la mediana:

$$\frac{16+19}{2}=\frac{35}{2}=17.5$$

La mediana de una distribución de frecuencias para datos agrupados

Primero se calcula $\frac{N}{2}$, a continuación determinamos cuál de las clases está a la mitad y le llamaremos clase de la media. Dentro de ella se localiza la mediana con una interpolación lineal en la forma siguiente:

Problema 7

a) Calcular la mediana de la distribución de frecuencias.

Clases	Frecuencia
28.5 – 33.5	7
33.5 – 38.5	13
38.5 – 43.5	20
43.5 – 48.5	11
48.5 – 53.5	5
Total	56

Solución:

Como $N = 56$

$$\frac{N}{2}=\frac{56}{2}=28$$

En este problema la clase de mediana es (38.5 – 43.5).

En la columna de las frecuencias y sumando 7 + 13 = 20 vemos que hay 20 frecuencias antes del valor de la clase de la mediana; los 8 que faltan se interpolan en el ancho de la clase de la mediana, que en este caso es de 5 (corresponde a la diferencia de 43.5 – 38.5).

Interpolamos con la relación proporcional (razones y proporciones) para obtener el valor de 8:

20 es a 5 como 1 es a x:

$$20:5::1:x$$
$$20x=5(1)$$
$$x=\frac{5}{20}$$

Como al 1 le corresponden $\frac{5}{20}$, para los 8 que faltan tenemos:

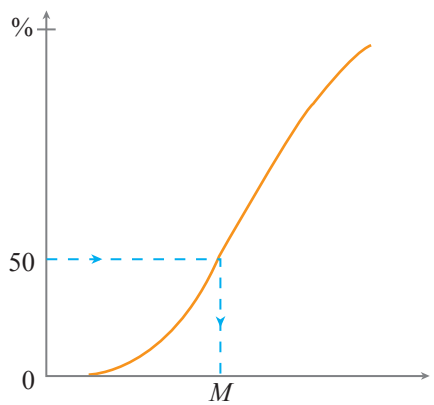
$$8\left(\frac{5}{20}\right)=\frac{40}{20}=2$$

Entonces $38.5+2=40.5$ es la mediana.

Representación gráfica de la mediana

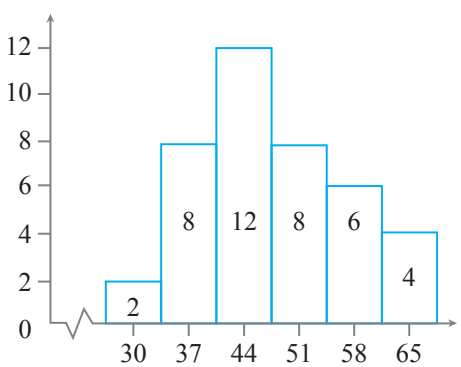
Solución:

La mediana se puede expresar utilizando la ojiva porcentual de una distribución, donde la mediana es la abscisa correspondiente al punto de ojiva cuya ordenada es el 50%, como lo hicimos en el capítulo anterior al tratar la ojiva proporcional, y como se indica a continuación.



Problema 8

Calcula la mediana en el histograma siguiente:



Solución:

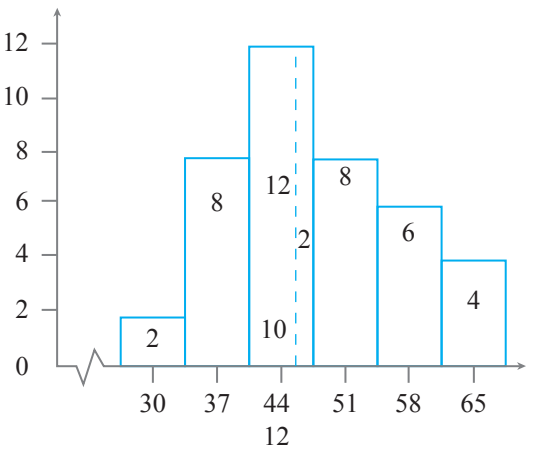
En un histograma, las áreas de ambos lados de la mediana deben ser iguales; por ello, sumamos y dividimos entre dos las áreas de los rectángulos.

$$2 + 8 + 12 + 8 + 6 + 4 = \frac{40}{2} = 20$$

Aceptamos que la mediana quedará dentro del rectángulo, cuya área es de 12(1) = 12 unidades, las cuales dividiremos en dos partes: una de 10 y la otra de 2 para que las áreas queden así:

El área de la izquierda con 2 + 8 + 10 = 20 unidades

El área de la derecha con 2 + 8 + 6 + 4 = 20 unidades



Numéricamente, los límites de la clase de la mediana son:

- Límite inferior $= \frac{37+44}{2} = \frac{81}{2} = 40.5$
- Límite superior $= \frac{44+51}{2} = \frac{95}{2} = 47.5$
- La amplitud de la clase es de 47.5 – 40.5 = 7
- La amplitud de 7 en 12 partes de $\frac{7}{12}$ que multiplicamos por $10 = \frac{70}{12} = 5.83$; entonces el lugar de la mediana es $40.5 + 5.83 = 46.33$

Moda

En un conjunto de datos de una distribución de frecuencias, la moda es el valor que ocurre con mayor frecuencia; por ejemplo, en los valores 1, 2, 5, 5, 6, 6, 6, 6, 7, 8, 9, 9, 9, la moda es el 6.

La moda es el valor más representativo o típico de una serie de valores, en el sentido que ocurre con mayor frecuencia.

Moda de datos agrupados

La moda en una distribución de datos agrupados es la marca del intervalo de clase que contiene la mayor frecuencia. La moda variará según la forma de agrupar.

Problema 9

Señala la moda de los valores siguientes sin agrupar y agrupados.

Valores sin agrupar

x	f
12	1
13	1
14	0
15	2
16	3
17	4
18	4
19	5
20	4
21	3
22	5
23	6
24	1
25	3
26	1
27	2
28	1
Total	46

Valores agrupados en 6 clases

Clases	Frecuencia
11.5 – 14.5	2
14.5 – 17.5	9
17.5 – 20.5	13
20.5 – 23.5	14
23.5 – 26.5	5
26.5 – 29.5	3
Total	46

En 4 clases

Clases	Frecuencia
11.5 – 16.5	7
16.5 – 21.5	20
21.5 – 26.5	16
26.5 – 31.5	3
Total	46

Solución:

La moda en los datos sin agrupar es 23, por corresponderle la mayor frecuencia, que es 6.

Para valores agrupados:

En el agrupamiento de 6 clases, la moda es 22, que es la marca de clase de 20.5–23.5, clase que contiene la mayor frecuencia **(14)**.

En el agrupamiento de 4 clases, la moda es 19, que es la marca de clase de 16.5–21.5, clase que contiene la mayor frecuencia **(20)**.

Uso de la media, la mediana y la moda

De las tres medidas de tendencia central, la media es la única que se puede trabajar, manipular o valorar algebraicamente; así, de la expresión de la media

$\bar{X} = \frac{\sum X}{N}$, despejando se tienen las relaciones:

$$\sum X = \bar{X}N$$

$$N = \frac{\sum X}{\bar{X}}$$

La media aritmética es sensible a los valores extremos, mientras que la mediana y la moda no lo son; por ejemplo, en la serie 4, 5, 7, 7, 9 la media es:

$$\overline{X} = \frac{\sum X}{N}$$

Sustituimos:

$$\overline{X} = \frac{4+5+7+7+9}{5} = \frac{32}{5} = 6.4$$

Si cambiamos el valor extremo 9 por 24, se tiene la serie 4, 5, 7, 7, 24, cuya media es:

$$\overline{X} = \frac{4+5+7+7+24}{5} = \frac{47}{5} = 9.4$$

El valor de la media varió de 6.4 a 9.4

La mediana y la moda no varían su valor a pesar del cambio.

La mediana y la moda de las series 4, 5, 7, 7, 9 y 4, 5, 7, 7, 24 es 7.

Debido a esta sensibilidad de la media a los valores extremos, a veces resulta que su valor produce efectos engañosos.

Problema 10

Para atender la demanda salarial de un grupo de ocho trabajadores se analiza su ingreso en pesos: 32, 40, 40, 45, 50, 55, 200, 300, respectivamente.

Solución:

$$\text{Media} = \overline{X} = \frac{32+40+40+45+50+55+200+300}{8} = \frac{762}{8}$$

$$= \$95.25$$

$$\text{Mediana} = \frac{45+50}{2} = \frac{95}{2}$$

$$= \$47.50$$

Moda = $40.00

Sólo dos personas obtienen ingresos altos y las seis restantes tienen salarios de $55.00 o menos; la media resultó engañosa. La mediana de $47.50 y la moda de $40.00 son más representativas.

Conclusiones:

1. Las tres medidas de tendencia central dan una apreciación de la distribución de los valores, pero si en un problema se debe hacer una apreciación con una sola medida, es mejor usar la mediana, la cual da el valor del medio.

2. Si la distribución es casi simétrica, se puede utilizar cualquiera de las tres medidas, cuyos resultados son aproximadamente iguales.

3. Si los datos no están ordenados, resulta más fácil el cálculo de la media que el de la mediana. La moda se localiza por la simple búsqueda del valor más frecuente.

4. En el uso de la mediana cuando los datos son irregulares y hay lagunas en los valores, su ubicación puede resultar falsa.

5. Si se quieren calcular totales, la única medida a utilizar es la media; por ejemplo, los consumos de agua, de combustible o de energía eléctrica para un periodo futuro.

6. Si se quieren ubicar las condiciones de una persona, la mediana es la medida más adecuada, ya que por comparación se observa si está por sobre la mitad o por debajo de ella.

7. Finalmente, la medida de tendencia central que se debe utilizar depende de la información disponible y el objeto que se desea alcanzar.

Media geométrica y armónica

La media, mediana y moda son las medidas de tendencia central más fáciles de calcular y las de mayor aplicación. Sin embargo, otras dos medidas de tendencia central se aplican en determinados problemas y, por ello, es conveniente conocerlas.

Media geométrica

Se define como la raíz n del producto de n términos. Su uso permite el cálculo de tasas de crecimiento.

$$\text{Media geométrica} = \sqrt[n]{X_1 X_2 \ldots X_n}$$

Problema 11

El crecimiento de las ventas de petróleo en los últimos cuatro años fue de 8, 16, 17 y 19%. Calcula la media geométrica anual de crecimiento.

Solución:

1.08 primer año

1.16 segundo año

1.17 tercer año

1.19 cuarto año

$$\text{Media geométrica} = \sqrt[4]{1.08(1.16)(1.17)(1.19)} = \sqrt[4]{1.744}$$

$$= 1.15$$

Para multiplicar 8, 16, 17 y 19%, a cada uno le agregamos un uno y al resultado le restamos uno por el razonamiento siguiente:

Si compras un bolígrafo en dos pesos y lo vendes en tres obtienes una ganancia de 50%; lo vendiste en $3.00 no en $0.50, de ahí que se agregue el uno.

Al resultado le quitamos el uno y lo multiplicamos por 100 para expresarlo en tanto por ciento.

Así:

$\sqrt[4]{1.742} = 1.15$ le quitamos el uno y queda 0.15 (100) = 15% es la media geométrica anual de crecimiento del petróleo. Si no agregamos el uno, tendríamos:

$$\sqrt[4]{0.08(0.16)(0.17)(0.19)} = \sqrt[4]{0.0004134}$$

El resultado está bien respecto a la multiplicación de números decimales pero no es el producto del 8, 16, 17 y del 19%.

Si tu calculadora no tiene capacidad para obtener raíces mayores a la raíz cuadrada, quizá tu profesor te permita dejar indicada la operación (la puedes calcular con logaritmos).

Problema 12

La población de un estado de la república creció en los últimos 5 años de 3 200 000 a 3 570 000 habitantes. Calcula la tasa de crecimiento total en los 5 años y la tasa de crecimiento anual.

Solución:

Número de habitantes que aumentaron:

$3\,570\,000 - 3\,200\,000 = 370\,000$

$$\text{Tasa de crecimiento en 5 años} = \frac{370\,000}{320\,000} = 0.1156 = 11.56\%$$

$$\text{Tasa de crecimiento anual } \frac{0.1156}{5} = 0.0231 = 2.31\%$$

Media armónica

La media armónica H de una serie de números es el recíproco de la media aritmética de los recíprocos de los números de la serie.

Sean los números $X_1, X_2, \ldots X_n$ la media armónica es $\dfrac{1}{H} = \dfrac{\frac{1}{X_1} + \frac{1}{X_2} \cdots + \frac{1}{X_n}}{N}$

Problema 13

Calcula la media armónica de los números 2, 5 y 7.

Solución:

$$\frac{1}{H} = \frac{\frac{1}{2} + \frac{1}{5} + \frac{1}{7}}{3} = \frac{\frac{35+14+10}{70}}{3} = \frac{\frac{59}{70}}{3} = \frac{59}{210}$$

$$\frac{1}{H} = \frac{59}{210}$$

$$H = \frac{210}{59} = 3.559 \text{ (se tomaron tres cifras decimales)}$$

La media armónica es 3.559.

Problema 14

Un obrero puede armar un motor en 3 días, mientras que otro obrero puede terminar el mismo trabajo en 4 días. Calcula el rendimiento de un trabajador representativo de los rendimientos de los dos obreros.

Solución:

$$\frac{1}{H}=\frac{\frac{1}{3}+\frac{1}{4}}{2}=\frac{\frac{4+3}{12}}{2}=\frac{\frac{7}{12}}{2}=\frac{7}{24}$$

$$\frac{1}{H}=\frac{7}{24}$$

$$H=\frac{24}{7}=3\frac{3}{7}$$

Rendimiento del trabajador representativo de $3\frac{3}{7}$.

Problema 15

Una compañía de transportes de carga tiene tres camiones diferentes que utiliza en el recorrido de *A* a *B*: 4, 3, 7 horas, respectivamente. Calcula el tiempo que empleará un camión para hacer el recorrido y que sirva de base para un estudio de costos.

Solución:

$$\frac{1}{H}=\frac{\frac{1}{X_1}+\frac{1}{X_2}+\cdots\frac{1}{X_n}}{N}$$

$$\frac{1}{H}=\frac{\frac{1}{4}+\frac{1}{3}+\frac{1}{7}}{3}=\frac{\frac{21+28+12}{84}}{3}=\frac{\frac{61}{84}}{3}=\frac{61}{252}$$

$$\frac{1}{H}=\frac{61}{252}$$

$H=\frac{252}{61}=4.131$ se tomaron tres cifras decimales

Tiempo: 4.131 (4 horas y casi 8 minutos).

La media armónica es 4.131.

Problema 16

Calcula la media geométrica y la media armónica de las siguientes series de números.

a) 3, 4, 5, 6, 7

b) 5, 6, 8, 9, 10, 11

Solución:

a) Media geométrica de 3, 4, 5, 6, 7

Sustituimos en $\sqrt[4]{X_1X_2...X_n}$

Media geométrica = $\sqrt[5]{3(4)(5)(6)(7)} = \sqrt[5]{2\,520} = 4.789$ se tomaron tres cifras decimales.

La media geométrica es 4.789

Media armónica de 3, 4, 5, 6, 7

Sustituimos en $\frac{1}{H} = \frac{\frac{1}{X_1} + \frac{1}{X_2} + \cdots \frac{1}{X_n}}{N}$

$$\frac{1}{H} = \frac{\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}}{5} = \frac{\frac{140 + 105 + 84 + 70 + 60}{420}}{5} = \frac{\frac{459}{420}}{5}$$

$$\frac{1}{H} = \frac{459}{2\,100}$$

$H = \frac{2\,100}{459} = 4.575$ (se tomaron tres cifras decimales)

La media armónica es 4.575.

b) Media geométrica de 5, 6, 8, 9, 10, 11

Sustituimos en $\sqrt[n]{X_1X_2 \ldots X_n}$

Media geométrica = $\sqrt[6]{5(6)(8)(9)(10)(11)} = \sqrt[6]{237\,600} = 7.869$

La media geométrica es 7.869

media armónica de 5, 6, 8, 9, 10, 11

Sustituimos en:

$$\frac{1}{H} = \frac{\frac{1}{X_1} + \frac{1}{X_2} + \cdots \frac{1}{X_n}}{N}$$

$$\frac{1}{H} = \frac{\frac{1}{5} + \frac{1}{6} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \frac{1}{11}}{6}$$

$$\frac{1}{H} = \frac{\frac{792 + 660 + 495 + 440 + 396 + 360}{3\,960}}{6}$$

$$\frac{1}{H}=\frac{\frac{3\,143}{3\,960}}{6}=\frac{3\,143}{23\,760}$$

$$\frac{1}{H}=\frac{3\,143}{23\,760}$$

$$H=\frac{23\,760}{3\,143}=7.559$$

La media armónica es 7.559.

Problema 17

Calcula la media aritmética, la mediana y la moda de las series de valores siguientes:

a) 2, 3, 7, 4, 7, 4, 1, 8

b) 5, 5, 3, 2, 2

c) 1, 9, 9, 9, 3, 0, 2, 2

Solución:

a) Serie 2, 3, 7, 4, 7, 4, 1, 8

Media aritmética

$$2+3+7+4+7+4+1+8=\frac{36}{8}=4.5$$

Mediana

1, 2, 3, 4, 4, 7, 7, 8

$$4+4=\frac{8}{2}=4$$

Moda

1, 2, 3, 4, 4, 7, 7, 9

Moda 4 y 7

b) Serie 5, 5, 3, 2, 2

Media aritmética

$$5+5+3+2+2=\frac{17}{5}=3.4$$

Media aritmética = 3.4

Mediana = 3

Moda 2 y 5

c) Serie 1, 9, 9, 9, 3, 0, 2, 2

Media aritmética

$$1+9+9+9+3+0+2+2=\frac{35}{8}=4.375$$

Mediana

0, 1, 2, 2, 3, 9, 9, 9

$$2+3=\frac{5}{2}=2.5$$

Mediana = 2.5

Moda = 9

Ejercicios de repaso

1. Las calificaciones de Juan en cinco materias fueron: 85, 77, 93, 76, 96. ¿Cuál es la media aritmética de sus calificaciones?

Sol. 85.4

2. Durante un examen médico los tiempos de reacción de una persona a determinados estímulos fueron 0.55, 0.42, 0.50, 0.48, 0.53, 0.50, 0.40 y 0.35 segundos respectivamente. Calcula el tiempo medio de reacción de la persona a los estímulos.

Sol. 0.47 *s*

3. ¿Cuál es la moda de los tiempos de reacción a determinados estímulos, citados en el problema anterior?

Sol. 0.50 *s*

4. Calcula la media aritmética de la distribución de frecuencias agrupadas de la tabla de frecuencias siguientes:

Intervalos	Marca de clase Marca X	Frecuencia (f)	$f(X)$
45 – 50	48	1	48
50 – 55	53	0	0
55 – 60	58	5	290
60 – 65	63	6	378
65 – 70	68	7	476
70 – 75	73	0	0
75 – 80	78	2	156

Sol. 64.19 (se tomaron 4 cifras decimales)

5. Los salarios mensuales de cinco trabajadores son:

2 500, 3 400, 1 800, 5 620, 4 200. Calcula la media aritmética. ***Sol.*** 3 504.00

6. Calcula la moda de las siguientes series de números.

a) 1, 2, 5, 6, 7, 6, 2, 6, 8, 9. ***Sol.*** 6

b) 5, 8, 3, 9, 7, 2, 1. ***Sol.*** No tiene moda

c) 7, 6, 8, 9, 8, 5, 4, 7. ***Sol.*** 7, 8 hay dos

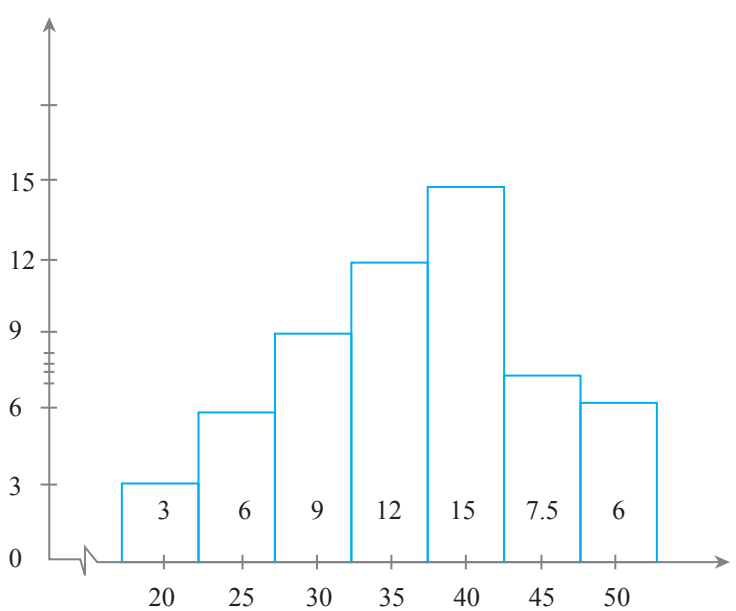
7. Calcula la mediana de la siguiente distribución de frecuencias:

Clases	Frecuencia
18.5 – 21.5	9
21.5 – 24.5	8
24.5 – 27.5	10
27.5 – 30.5	12
30.5 – 33.5	8
33.5 – 36.5	7
36.5 – 39.5	5

Sol. 28.125

8. Calcula el valor de la media en el histograma siguiente, y trázala en su lugar.

Sol. 36 (no se tomaron dos cifras decimales)



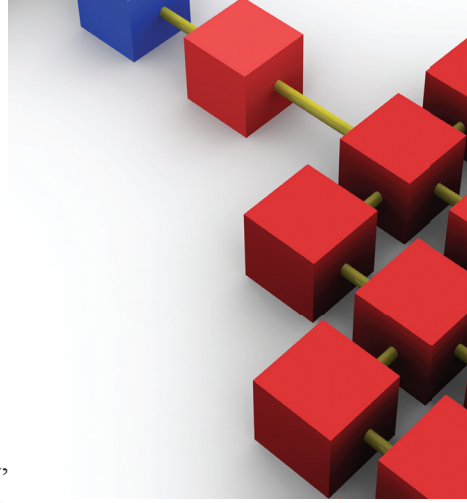
9. Identifica la moda en los valores sin agrupar siguientes:

X	f
28	2
29	3
30	0
31	5
32	4
33	1
34	0
35	6
36	4
Total	25

Sol. 35

10. Calcula la media armónica de los números 8,12,15,18.

Sol. 12.10 (se tomaron dos cifras decimales)



Capítulo 10

Medidas de dispersión

Generalidades

En el capítulo anterior estudiamos las medidas de tendencia central: media aritmética, mediana y moda; asimismo, analizamos la media geométrica y la media armónica, que describen el comportamiento de los datos en una distribución de frecuencias.

Conviene aclarar que estas medidas no proporcionan información sobre la forma en que están distribuidos o dispersos los valores con relación a la tendencia central y poco informan sobre un dato específico con relación a los otros en la distribución de frecuencias.

En este capítulo estudiaremos el rango, la desviación cuartil, la desviación media, la varianza y la desviación estándar. Todos estos recursos nos ayudarán a medir la dispersión.

Conceptos clave

- Rango
- Cuartil
- Decil
- Rango intercuartil
- Rango semintercuartil o desviación cuartil
- Desviación media
- Varianza

Problema 1

En una escuela se aplicó un examen extraordinario para los 40 alumnos que reprobaron matemáticas y física. Los estudiantes fueron calificados sobre 30 puntos y obtuvieron las calificaciones que se expresan en el siguiente cuadro de frecuencias agrupadas. Juan obtuvo 16 puntos en los dos exámenes que presentó, calcula qué resultado debe esperar en su calificación.

Clases (calificaciones)	Frecuencias	
	Matemáticas	Física
0.5 – 3.5	2	3
3.5 – 6.5	4	3
6.5 – 9.5	9	0
9.5 – 12.5	10	1
12.5 – 15.5	8	2
15.5 – 18.5	4	2
18.5 – 21.5	0	7
21.5 – 24.5	2	9
24.5 – 27.5	1	12
27.5 – 30.5	0	1
	40	40

Solución:

Juan obtuvo 16 puntos en ambos exámenes. En matemáticas su calificación será bastante alta porque sólo hay 3 calificaciones mejores, pero el resultado de su examen de física no es bueno porque hay 29 mejores.

Para la interpretación de los resultados individuales de estos exámenes se necesita más información que permita apreciar la dispersión de los valores en el entorno de la tendencia central.

Rango

En toda distribución hay valores extremos, uno menor y otro mayor; la diferencia entre estos valores se llama **rango** y en él están distribuidos todos los demás valores, por eso también se le llama *recorrido*.

El rango es una medida de dispersión y es la más fácil de obtener; sin embargo, su uso es limitado porque es muy influenciable por la presencia de valores extremos de poca frecuencia. Se piensa que cuanto mayor es el rango, mayor es la dispersión de los datos, lo cual conduce a apreciaciones falsas.

Problema 2

Considera las frecuencias *A* y *B*. Calcula la media aritmética de ambas.

Solución:

X	Frecuencias	
	A	*B*
55	0	1
54	1	0
53	0	2
52	0	0
51	2	0
50	3	3
49	9	7
48	17	19
47	3	2
46	3	4
45	1	1
	39	39

$$\text{Rango} = 55 - 45 = 10$$

Cálculo de las medias aritméticas

Para *A:*

$$\sum_{i=1}^{n} f_i X_i = 55(0) + 54(1) + 53(0) + 52(0) + 51(2) + 50(3)$$
$$+ 49(9) + 48(17) + 47(3) + 46(3) + 45(1)$$
$$= 0 + 54 + 0 + 0 + 102 + 150 + 441 + 816 + 141 + 138 + 45 = 1\,887$$

$$\overline{X} = \frac{\sum f X}{N} = \frac{1\,887}{39} = 48.38 \text{ (se tomaron dos cifras decimales)}$$

Para *B:*

Obtenemos $\sum_{i=1}^{n} f_i X_i = 55(1)+54(0)+53(2)+52(0)+51(0)+50(3)$

$+49(7)+48(19)+47(2)+46(4)+45(1)$

$=55+0+106+0+0+150+343+912+94$

$+184+45=1\,889$

$\bar{X}=\frac{\sum fX}{N}=\frac{1\,889}{39}=48.43$ (se tomaron dos cifras decimales)

El rango en ambas es el mismo y la diferencia en la media aritmética de sus distribuciones es de 5 centésimas. Sin embargo, sus distribuciones tendrán una diferencia mayor si en la distribución *A* suprimimos 54 de frecuencia 1, entonces queda más compacta ya que:

Rango = 51 − 45 = 6

Y la media:

$\bar{X}=\frac{\sum fX}{N}=\frac{1\,833}{38}=48.23$

Observa que el resultado varió al suprimir un dato en uno de los extremos de la frecuencia *A*. ahora tenemos:

Rango = 6

Media = 48.23

Cuartiles y deciles

Para conocer los intervalos dentro de los cuales quedan representados proporcionalmente los términos de una distribución, se divide la distribución de frecuencias en cuatro partes iguales, cada una contiene igual número de observaciones (25% del total).

Los puntos de separación de los valores de *X* se llaman **cuartiles**. El primer cuartil corresponde al 25% y se designa con Q_1. El segundo se designa con Q_2, que representa el valor de 50% y coincide con la mediana. El tercer cuartil es Q_3 y representa 75% de las observaciones que están por debajo de él.

Si en lugar de dividir en cuatro partes iguales se hace en diez, se tienen nueve puntos de división, correspondiendo a cada punto un **decil**, de donde, el primer decil es el valor por debajo del cual está el 10% de las observaciones, para el segundo decil el 20%, y así sucesivamente.

Cálculo de los cuartiles

Problema 3

Calcula los cuartiles en el cuadro de frecuencias agrupadas, en donde se han registrado las alturas de un grupo de alumnos.

Solución:

Dividimos el total *N* de las frecuencias acumuladas entre cuatro y obtenemos el número de observaciones que hay en el primer cuartil.

Clases (estaturas)	Frecuencias (alumnos)
121.5 – 126.5	2
126.5 – 131.5	3
131.5 – 136.5	8
136.5 – 141.5	23
141.5 – 146.5	27
146.5 – 151.5	20
151.5 – 156.5	16
156.5 – 161.5	3
161.5 – 166.5	2
Total	$N = 104$

$\frac{N}{4} = \frac{104}{4} = 26$ el primer cuartil cae en la clase 136.5 – 141.5.

Las tres primeras clases contienen 13 alumnos (sumamos 2 + 3 + 8 = 13). Para los 13 que faltan, los calculamos por interpolación lineal:

$23 : 5 :: 1 : x$

El 5 número se obtiene de 141.5 – 136.5 = 5

$23x = 5$

$$x = \frac{5}{23}$$

Como al número 1 le corresponde $\frac{5}{13}$, para los 13 que faltan, tenemos:

$$13\left(\frac{5}{23}\right) = \frac{65}{23} = 2.8 \quad \text{(se redondeó a una cifra decimal)}$$

Entonces 136.5 + 2.8 = 139.3 Éste es el valor del primer cuartil.

Segundo cuartil

De la clase 136.5 – 141.5 que tiene una frecuencia de 23, se tomaron 13 para el primer cuartil, entonces quedan 10 para el segundo, para 26 faltan 16, que caen en la clase 141.5 – 146.5; éstos 16 los calcularemos por intepolación polar:

$27 : 5 :: 1 : x$

$27x = 5$

$$x = \frac{5}{27}$$

Como al número 1 le corresponde $\frac{5}{27}$, para los 16 que faltan, tenemos:

$$16\left(\frac{5}{27}\right) = \frac{80}{27} = 3 \quad \text{(se redondeó a una cifra decimal)}$$

Entonces, 141.5 + 3 = 144.5; por lo tanto, éste es el valor del segundo cuartil y significa que 52 alumnos tienen una estatura menor o igual a 144.5 cm, que es la mediana de las estaturas.

Tercer cuartil

De la clase 141.5 – 146.5 que tiene una frecuencia de 27, se tomaron 16 para el segundo cuartil, entonces quedan 11 para el tercero, para 26 faltan 15, que caen en la clase 146.5 – 151.5; éstos 15 los calcularemos por interpolación polar.

$20 : 5 :: 1 : x$

$20x = 5$

$$x = \frac{5}{20}$$

Como al número 1 le corresponde $\frac{1}{4}$, para los 15 que faltan tenemos:

$$15\left(\frac{1}{4}\right)=\frac{15}{4}=3.7$$

Entonces, $146.5 + 3.7 = 150.2$ Es el valor del tercer cuartil.

Rango intercuartil

Es el resultado de la diferencia entre el tercer cuartil Q_3 y el primero Q_1. Se expresa así:

$$\text{Rango intercuartil} = Q = Q_3 - Q_1$$

Si después de haber aplicado la media aritmética se quiere evitar la influencia de los valores extremos, se analiza únicamente la situación intermedia de la distribución de frecuencias aplicando el rango intercuartil.

El **rango semintercuartil** o **desviación cuartil** es la mitad del rango intercuartil; se representa con Q_D

$$Q_D = \frac{Q_3 - Q_1}{2}$$

El rango semicuartil (desviación cuartil) mide la dispersión con mayor precisión que el rango; sin embargo, presenta las limitaciones siguientes:

a) No toma en consideración todos los valores de la distribución de frecuencias y puede suceder que los valores menores a Q_1 o superiores a Q_3 estén muy compactos o muy dispersos, y el valor de Q sería el mismo.

b) No es posible, conociendo únicamente Q, hacer la ubicación precisa de una observación dentro de la distribución de frecuencias.

c) Igual que la mediana, no tiene propiedades que permitan su uso en las relaciones matemáticas que utiliza la estadística.

Problema 4

Calcula el rango intercuartil y la desviación cuartil de la distribución del cuadro anterior, en que se manejaron las estaturas de 104 alumnos y se calcularon $Q_1 = 139.3$ y $Q_3 = 150.2$, que significa: 18 alumnos miden menos de 139.3 cm y 52 miden menos de 150.2 cm.

Solución:

$$\text{Rango intercuartil} = Q = Q_3 - Q_1$$

$$Q = 150.2 - 139.3 = 10.9$$

$$\text{Desviación cuartil} = Q_D = \frac{Q_3 - Q_1}{2}$$

$$Q_D = \frac{10.9}{2} = 5.45$$

Desviación media y varianza

Son medidas de dispersión que tienen relación con la media aritmética porque las tres tienen propiedades algebraicas que les permiten su uso en relaciones matemá-

ticas, que son la base estructural de los análisis estadísticos. Por sus propiedades algebraicas, son las medidas de dispersión de mayor uso e importancia.

Desviación media

Una de las propiedades de la media aritmética establece que en toda distribución la suma de las desviaciones de cada valor de la variable respecto a la media es cero; esto es, que la suma de las desviaciones $(X - \bar{X})$ de las variables mayores que la media es igual y de signo opuesto a la suma de las desviaciones de las variables menores que la media.

Por lo tanto, para obtener las desviaciones medias se usan los valores absolutos de las desviaciones, que se expresa $|X - \bar{X}|$. Este valor corresponde al valor positivo de $X - \bar{X}$, no importando que X sea positivo o negativo, por ejemplo: $|5| = 5$, $|-5| = 5$, si $X = 0$, entonces $|X| = 0$.

La media aritmética de los valores absolutos de las desviaciones de cada uno de los valores de la variable, respecto a la media aritmética, es la desviación media.

La desviación media es una medida de dispersión muy objetiva y cuanto mayor sea su valor, mayor es la dispersión de los datos. Sin embargo, no proporciona una relación matemática precisa entre su magnitud y la posición de un dato dentro de la distribución; además, al tomarse los valores absolutos, mide la desviación de una observación sin mostrar si está por encima o por debajo de la media aritmética.

Se expresa:

$$\text{Desviación media} = DM = \frac{\sum |X - \bar{X}|}{N}$$

y para la distribución de frecuencias agrupadas:

$$DM = \frac{\sum f |X - \bar{X}|}{N}$$

Problema 5

Calcula la DM de los números: 6, 3, 4, 12, 10, 2, 7, 5.

Solución:

Primero calculamos el valor de $\bar{X}$ (media aritmética):

$$\bar{X} = \frac{\sum X}{N} = \frac{6+3+4+12+10+2+7+5}{8} = \frac{49}{8} = 6.12$$

Ahora, obtenemos la desviación media:

$$DM = \frac{\sum |X - \bar{X}|}{N} = \frac{|6-6.12|+|3-6.12|+|4-6.12|+|12-6.12|+}{8}$$
$$\underline{|10-6.12|+|2-6.12|+|7-6.12|+|5-6.12|}$$

Tomamos los valores absolutos:

$$DM = \frac{0.12+3.12+2.12+5.88+3.88+4.12+0.88+1.12}{8} = \frac{21.24}{8}$$

$= 2.655$ es la desviación media

Problema 6

Calcula la desviación media de la distribución de frecuencias agrupadas que citamos en el problema 3. Recuerda que en ese problema calculamos los cuartiles.

Clases (estatura)	Frecuencias (alumnos)
121.5 – 126.5	2
126.5 – 131.5	3
131.5 – 136.5	8
136.5 – 141.5	23
141.5 – 146.5	27
146.5 – 151.5	20
151.5 – 156.5	16
156.5 – 161.5	3
161.5 – 166.5	2

Solución:

Disponemos los datos así:

Clases (intervalos)	Marca	f	fX	$\lvert X-\bar{X}\rvert$	$f\lvert X-\bar{X}\rvert$
121.5 – 126.5	124	2	248	20.62	41.24
126.5 – 131.5	129	3	387	15.62	46.86
131.5 – 136.5	134	8	1 072	10.62	84.96
136.5 – 141.5	139	23	3 197	5.62	129.26
141.5 – 146.5	144	27	3 888	0.62	16.74
146.5 – 151.5	149	20	2 980	4.38	87.60
151.5 – 156.5	154	16	2 464	9.38	150.08
156.5 – 161.5	159	3	477	14.38	43.14
161.5 – 166.5	164	2	328	19.38	38.76
Totales		104	15 041		638.64

Los datos numéricos para llenar el cuadro y calcular la desviación media (DM) se obtuvieron así:

a) Las clases también se suelen citar como intervalos.

b) La marca o *marca de clase* es el punto medio entre los extremos de un intervalo, en el ejemplo son:

$$\frac{121.5+126.5}{2}=\frac{248.0}{2}=124$$

$$\frac{126.5+131.5}{2}=\frac{258.0}{2}=129$$

Y así se calculan las demás.

c) Frecuencia (f) es el número de elementos que hay en el intervalo, se tomaron del cuadro donde se organizaron las estaturas de los alumnos en forma ascendente.

d) $f(x)$ es el resultado del producto de la marca por la frecuencia.

124 (2) – 248

129 (3) – 387

Y así se calcularon las demás.

e) $|X - \bar{X}|$ es el valor absoluto de la diferencia de X y $\bar{X}$.

El valor de X está dentro del intervalo correspondiente, por eso tomamos el de la marca que lo representa. El de $\bar{X}$ es:

$$\bar{X} = \frac{\sum f X}{N} = \frac{15\,041}{104} = 144.62$$

Operaciones para obtener los resultados de $|X - \bar{X}|$

124 – 144.62	= 20.62
129 – 144.62	= 15.62
134 – 144.62	= 10.62
139 – 144.62	= 5.62
144 – 144.62	= 0.62
149 – 144.62	= 4.38
154 – 144.62	= 9.38
159 – 144.62	= 14.38
164 – 144.62	= 19.38

f) Hechos los cálculos necesarios para obtener los valores del cuadro, tenemos:

Desviación media $DM = \dfrac{\sum f\,|X - \bar{X}|}{N}$

$$DM = \frac{638.64}{104}$$

$DM = 6.14$ (se tomaron 2 cifras decimales)

Varianza

Es la media aritmética de los cuadrados de desviaciones respecto a la media aritmética. Se representa s^2.

Al calcular la desviación media fue necesario ignorar los signos negativos y tomar los valores absolutos de las desviaciones respecto a la media aritmética.

Si elevamos al cuadrado las desviaciones, todas las desviaciones dan positivas, sumando los cuadrados de las desviaciones y dividiendo entre N, obtenemos la varianza.

La varianza sirve de base para calcular la desviación estándar o desviación típica, que es la más importante de todas las medidas de dispersión.

La varianza para datos no agrupados se obtiene con:

$$s^2 = \frac{\sum \left(X - \bar{X} \right)^2}{N}$$

Para datos agrupados:

$$s^2 = \frac{\sum f \left(X - \bar{X} \right)^2}{N}$$

Nota: Ahora usamos paréntesis y no barras para denotar el valor absoluto.

Si en las relaciones anteriores sustituimos $X - \bar{X}$ con x, se tiene:

$s^2 = \dfrac{\sum x^2}{N}$ para datos no agrupados

$s^2 = \dfrac{\sum fx^2}{N}$ para datos agrupados

A continuación desarrollaremos otra relación (fórmula) que permite también el cálculo de la varianza.

Desarrollamos el binomio:

$$\sum \left(X - \bar{X} \right)^2 = \sum X^2 - 2 \sum X \bar{X} + \sum \bar{X}^2 = \sum X^2 - 2 \bar{X} \sum X + \bar{X} \sum \bar{X}$$

Sustituimos el valor de la media aritmética donde se cita $\bar{X}$:

$$= \sum X^2 - 2 \frac{\sum X}{N} \sum X + \frac{\sum X}{\cancel{N}} \cancel{N} X$$

$$= \sum X^2 - 2 \frac{\sum X^2}{N} + \sum X \frac{\sum X}{N}$$

$$= \sum X^2 - \frac{\left(\sum X \right)^2}{N}$$

Ahora sustituimos en $s^2 = \dfrac{\sum \left(X - \bar{X} \right)^2}{N}$

Por lo tanto:

$$s^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N} \right)^2$$

$s^2 = \dfrac{\sum X^2}{N} - \bar{X}^2$ para datos no agrupados

$s^2 = \dfrac{\sum f X^2}{M} - \bar{X}^2$ para datos agrupados

Problema 7

Calcula la desviación media *DM* y la varianza de la serie de números 9, 11, 1, 8, 14, 5, 6, 7, 11, 9.

Solución:

Calculamos la media aritmética $\bar{X}$

$$\bar{X} = \frac{\sum X}{N} = \frac{9+11+1+8+14+5+6+7+11+9}{10} = \frac{81}{10}$$

$$\bar{X} = 8.1$$

Para obtener *DM* tomamos el valor absoluto

$$DM = \frac{\sum |X - \bar{X}|}{N}$$

$$= \frac{|9-8.1|+|11-8.1|+|1-8.1|+|8-8.1|+|14-8.1|+|5-8.1|+}{10}$$

$$\underline{|6-8.1|+|7-8.1|+|11-8.1|+|9-8.1|}$$

$$= \frac{0.9+2.9+7.1+0.1+5.9+3.1+2.1+1.1+2.9+0.9}{10} = \frac{27}{10}$$

$$DM = 2$$

Obtenemos la varianza y como $\bar{X} = 8.1$, sustituimos en:

$$s^2 = \frac{\sum \left(X - \bar{X}\right)^2}{N}$$

$$= \frac{(9-8.1)^2+(11-8.1)^2+(1-8.1)^2+(8-8.1)^2+(14-8.1)^2+(5-8.1)^2+}{10}$$

$$\underline{(6-8\ 1)^2+(7-8.1)^2+(11-8.1)^2+(9-8.1)^2}$$

$$= \frac{(0.9)^2+(2.9)^2+(7.1)^2+(0.1)^2+(5.9)^2+(3.1)^2+(2.1)^2+(1.1)^2+}{10}$$

$$\underline{(2.9)^2+(0.9)^2}$$

$$= \frac{0.81+8.41+50.41+0.01+34.81+9.61+4.41+1.21+8.41+0.81}{10}$$

$$= \frac{118.90}{10}$$

$$s^2 = 11.89$$

Problema 8

Retomamos el problema 6 en que calculamos la desviación media de una distribución de frecuencia acumulada. Agregamos las columnas $(X - \bar{X})^2$ y $f(X - \bar{X})^2$ para obtener la varianza y expresar su valor.

Solución:

Primero citamos los datos obtenidos en el problema 6 y agregamos las columnas $(X-\bar{X})^2$ y $f(X-\bar{X})^2$ para determinar el resultado.

Clases (intervalos)	Marca	f	fx	$\lvert X-\bar{X}\rvert$	$f\lvert X-\bar{X}\rvert$	$(X-\bar{X})^2$	$f(X-\bar{X})^2$
121.5 – 126.5	124	2	248	20.62	41.24	425.18	850.36
126.5 – 131.5	129	3	387	15.62	46.86	243.98	731.94
131.5 – 136.5	134	8	1072	10.62	84.96	112.78	902.24
136.5 – 141.5	139	23	3 197	5.62	129.26	31.58	726.34
141.5 – 146.5	144	27	3 888	0.62	16.74	0.38	10.26
146.5 – 151.5	149	20	2 980	4.38	87.60	19.18	383.60
151.5 – 156.5	154	16	2 464	9.38	150.08	87.98	1 407.68
156.5 – 161.5	159	3	477	14.38	43.14	206.78	620.34
161.5 – 166.5	164	2	328	19.38	38.76	375.58	751.16
Totales		104	15 041		638.64		6 383.92

Analiza el procedimiento que seguimos para completar el cuadro:

- Al referirnos al cuadro que se elaboró para calcular la desviación media indicamos cómo obtuvimos la marca, la frecuencia (f), $f(X)$, que ese producto de la marca por la frecuencia, el valor absoluto de $|X-\bar{X}|$ y el producto de la frecuencia por este valor absoluto.
- Para obtener $(X-\bar{X})^2$ elevamos al cuadrado los valores obtenidos en la columna $|X-\bar{X}|$.
- Los resultados de $(X-\bar{X})^2$ los multiplicamos por los datos de f, y obtenemos $f(X-\bar{X})^2$.

La varianza la calculamos sustituyendo valores en las fórmulas:

$$\bar{X}=\frac{\Sigma fX}{N}=\frac{15\,041}{104}=144.62 \quad \text{(se tomaron dos cifras decimales)}$$

$$s^2=\frac{\Sigma f\left(X-\bar{X}\right)^2}{N}=\frac{6\,383.92}{104}$$

$$s^2=61.38 \quad \text{(se tomaron dos cifras decimales)}$$

Ejercicios de repaso

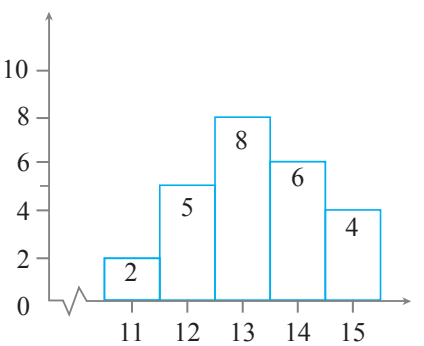
1. Calcula la desviación media de 1, 4, 12, 10, 7, 2.

Sol. 3. 66 (se tomaron dos cifras decimales)

2. Calcula la media geométrica de 5, 3, 6, 7, 6, 12, 10.

Sol. 6.43 (se tomaron dos cifras decimales)

3. Calcula el valor de la media aritmética en el histograma siguiente y trázala en su lugar.



Sol. 13 (no se tomó la cifra decimal)

4. Calcula la desviación media de la frecuencia agrupada siguiente:

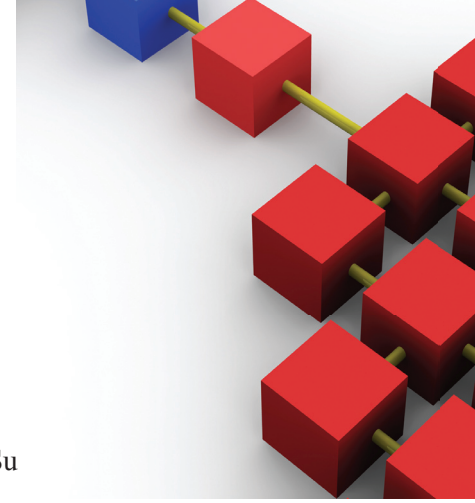
Clase	Frecuencias
151.5 – 156.5	3
156.5 – 161.5	8
161.5 – 166.5	12
166.5 – 171.5	10
171.5 – 176.5	6

Sol. 4.79 (se tomaron dos cifras decimales)

5. Con los datos obtenidos para calcular la desviación media de la misma frecuencia agrupada, citada en el problema anterior, determina el valor de la varianza.

Clase	Frecuencias
151.5 – 156.5	3
156.5 – 161.5	8
161.5 – 166.5	12
166.5 – 171.5	10
171.5 – 176.5	6

Sol. 33.56 (se tomaron 2 cifras decimales)



Capítulo 11

Desviación estándar o típica

Definición

La desviación estándar o desviación típica es la raíz cuadrada de la varianza. Su fórmula es:

$$s = \sqrt{s^2}$$

a) Para datos no agrupados:

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

Si escribimos $x = X - \bar{X}$, tenemos:

$$s = \sqrt{\frac{\sum x^2}{N}}$$

Si nos referimos a la fórmula modificada, estudiada en la varianza, tenemos:

$$s = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2}$$

$$s = \sqrt{\frac{\sum X^2}{N} - \bar{X}^2}$$

b) Para datos agrupados:

$$s = \sqrt{\frac{\sum f(\bar{X})^2}{N}}$$

$$s = \sqrt{\frac{\sum f X^2}{N} - \left(\frac{\sum f X}{N}\right)^2}$$

$$s = \sqrt{\frac{\sum f X^2}{N} - X^2}$$

La desviación estándar es la más importante de todas las medidas de dispersión porque incluye más o menos 68% de los términos de una distribución normal; además, por sus propiedades algebraicas se utiliza con facilidad en el análisis estadístico.

Hasta este punto hemos analizado diferentes medidas de dispersión: rango, *desviación media* y *varianza*. Sin embargo, los resultados que podemos obtener del rango son pocos porque representa una vaga apreciación de la amplitud de los datos. La diferencia con el resto de las medidas se valora respecto a la cantidad de datos que se incluye en sus aplicaciones.

En una distribución normal, el rango intercuartil corresponde a 50%.

La desviación media incluye aproximadamente 58%.

Tomando en cuenta que no podemos asegurar la completa efectividad de las herramientas de análisis, el resto de los porcentajes lo consideraremos como un riesgo calculado.

Problema 1

Calcula la varianza en la serie de números 9, 11, 1, 8, 14, 5 y a continuación, la desviación estándar entre ellos.

Solución:

Primero calculamos la media aritmética:

$$\overline{X} = \frac{\sum X}{N} = \frac{9+11+1+8+14+5}{6} = \frac{48}{6} = 8$$

$$\overline{X} = 8$$

Calculamos a continuación la varianza:

$$s^2 = \frac{\sum (X - \overline{X})^2}{N}$$

$$= \frac{(9-8)^2 + (11-8)^2 + (1-8)^2 + (8-8)^2 + (14-8)^2 + (5-8)^2}{6}$$

$$= \frac{1^2 + 3^2 + (-7)^2 + 0^2 + 6^2 + (-3)^2}{6} = \frac{1+9+49+0+36+9}{6} = \frac{104}{6}$$

$$= 17.33 \text{ (se tomaron dos cifras decimales)}$$

Si para obtener la varianza aplicamos, con los mismos datos, la siguiente fórmula:

$$s^2 = \frac{\sum X^2}{N} - \overline{X}^2$$

$$= \frac{9^2 + 11^2 + 1^2 + 8^2 + 14^2 + 5^2}{6} - 8^2$$

$$= \frac{81+121+1+64+196+25}{6} - 64$$

$$= \frac{488}{6} - 64 = 81.33 - 64$$

$s^2 = 17.33$ Obtuvimos los mismos resultados.

Ahora obtenemos la desviación estándar:

$$s^2 = 17.33$$

$$s = \sqrt{17.23} = 4.16$$

Problema 2

Calcula la desviación estándar de la distribución de frecuencias agrupadas que se cita. Aplica la siguiente fórmula:

$$s = \sqrt{\frac{\sum f X^2}{N} - \left(\frac{\sum f X}{N}\right)^2}$$

Clases (intervalos)	Marca	f
37 – 39	38	2
39 – 41	40	10
41 – 43	42	13
43 – 45	44	32
45 – 47	46	14
47 – 49	48	7
49 – 51	50	3
51 – 53	52	2
53 – 55	54	1

Solución:

Desarrollamos un cuadro para mostrar los datos que permitirán aplicar la fórmula.

Clases (intervalos)	Marca (intervalos)	f	fX	X^2	fX^2
37 – 39	38	2	76	1 444	2888
39 – 41	40	10	400	1 600	16 000
41 – 43	42	13	546	1 764	22 932
43 – 45	44	32	1 408	1 936	61 952
45 – 47	46	14	644	2 116	29 624
47 – 49	48	7	336	2 304	16 128
49 – 51	50	3	150	2 500	7 500
51 – 53	52	2	104	2 704	5 408
53 – 55	54	1	54	2 916	2 916
Totales		84	3 718	19 284	165 348

Aplicamos la fórmula:

$$\frac{\sum f X^2}{N} = \frac{165\,348}{84} = 1\,968.42 \text{ (se tomaron dos cifras decimales)}$$

$$\left(\frac{\sum f X}{N}\right)^2 = \left(\frac{3\,718}{84}\right)^2 = 1\,958.94 \text{ (se tomaron dos cifras decimales)}$$

Sustituimos en:

$$s = \sqrt{\frac{\sum f X^2}{N} - \left(\frac{\sum f X}{N}\right)^2}$$

$$s = \sqrt{1\,968.42 - 1\,958.94} = \sqrt{9.48}$$

$$s = 3.07$$

Dispersión relativa. Coeficientes de variación

Para realizar comparaciones entre series de observaciones se aplica la dispersión relativa, que es adimensional.

$$\text{Dispersión relativa} = \frac{\text{Dispersión absoluta}}{\text{media}}$$

Si la dispersión es la desviación estándar s, entonces la dispersión relativa recibe el nombre de *coeficiente de variación* y se representa así: $V = \frac{s}{\bar{X}}$.

Las medidas de dispersión (rango, desviación cuartil, desviación media, varianza y desviación estándar) son medidas de dispersión absoluta y se expresan en las mismas unidades con las que se mide la variable.

No es posible realizar una comparación entre dos o más series de observaciones si aplicamos la dispersión absoluta. Se pueden presentar dos posibles casos:

- Primero

  Las unidades de medida de las observaciones son iguales.

  En este caso, las dos series pueden tener medias aritméticas diferentes y por estar expresadas en las mismas unidades, las desviaciones estándar son comparables, pero no aportan una apreciación correcta sobre las series que se comparan.

- Segundo

  Las unidades de medida de las observaciones son desiguales.

  Entonces no es posible la comparación por medio de la dispersión absoluta, por ejemplo, si se trata de comparar calificaciones que tienen diferentes rangos o comparar una serie de precios en pesos y otra en dólares.

Problema 3

La media aritmética de los salarios de una empresa automotriz es de 14 600 y su desviación estándar es de 147. En una empresa maquiladora en la que se fabrican carburadores para automóviles la media aritmética de los salarios es de 5 700 y su desviación estándar es de 38. Calcula las dos series y decide cuál de las dos tiene mayor variación de salarios.

Solución:

Industria automotriz:

Coeficiente de variación $V = \frac{s}{\overline{X}} = \frac{147}{14\,600} = 0.010 = 1\%$

Industria maquiladora:

$$V = \frac{38}{5700} = 0.006 = 0.6\%$$

La industria automotriz tiene mayor variación en los salarios.

Ejercicios de repaso

1. Calcula la varianza de los números 3, 2, 1, 7, 5, así como su desviación estándar.

Sol. Varianza 4.64
Desviación estándar 2.15

2. Calcula la desviación media de los valores 7, 8, 9, 5, 4.

Sol. 1.68

3. Identifica la moda en los siguientes valores sin agrupar.

X	*f*
15	3
16	0
17	7
18	4
19	2
20	0
21	5
Total	21

Sol. 17

4. Calcula la moda de las series siguientes:

a. 3, 2, 4, 8, 9, 10, 1, 5

Sol. No hay

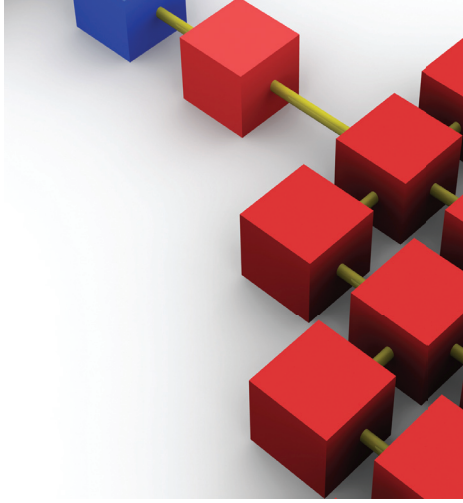
b. 4, 5, 6, 5, 1, 8, 9, 8, 3, 1, 3

Sol. Hay tres
3, 5 y 8

5. Calcula la desviación estándar aplicando la fórmula $s = \sqrt{\frac{\Sigma f X^2}{N} - \left(\frac{\Sigma f X}{N}\right)^2}$ en la distribución de frecuencias agrupadas siguiente.

Clase	Marca	Frecuencias
53 – 58	55.5	8
58 – 63	60.5	12
63 – 68	65.5	15
68 – 73	70.5	14
73 – 78	75.5	23

Sol. 6.88



Capítulo 12

Distribución de probabilidades discretas

Binomial o de Bernoulli

De Poisson

Conceptos clave

- Distribución binomial
- Distribución de Bernoulli
- Ensayos binomiales o de Bernoulli
- Probabilidad de éxito
- Probabilidad de fracaso
- Probabilidades acumulativas

Binomial

A. La distribución binomial se cita frecuentemente como **distribución de Bernoulli** en honor al matemático suizo Jacobo Bernoulli, quien la dedujo a finales del siglo XVII.

Se denominan **ensayos binomiales** o **ensayos de Bernoulli** a aquellos que tienen las características siguientes:

a) En una ejecución cualquiera hay exactamente dos resultados posibles: un éxito o un fracaso.

b) Hay n ejecuciones, donde n es un número entero positivo fijado de antemano.

c) La probabilidad de éxito para todas las ejecuciones es la misma.

d) Todas las ejecuciones son independientes.

Si consideramos el experimento de lanzar una moneda al aire, cada intento es un ensayo y tiene dos resultados posibles: éxito E o fracaso F.

Es posible que al realizar este juego no se fije de antemano el número de ejecuciones y aunque no cumple la condición B, se considera como un ensayo de Bernoulli porque lanzar monedas al aire sólo puede realizarse un número finito de veces.

Así como en este ejemplo, en otros muchos casos prácticos se obtienen resultados útiles, suponiendo que el experimento satisface las cuatro condiciones citadas o bien, sólo cumple parte de ellas. Analiza el caso siguiente:

Problema 1

En una fábrica de computadoras se elige una muestra de $n = 30$ para clasificar los productos elaborados como no defectuosos (éxito) y los defectuosos (fracaso).

Si durante el proceso de fabricación se realizan algunas modificaciones en la maquinaria o hay cambio de turno, podríamos asegurar, teóricamente, que la probabilidad P no es la misma para toda la muestra; sin embargo, consideramos que el cambio es tan pequeño que P no varía. Así, podemos considerar el problema como una serie de ensayos de Bernoulli.

Problema 2

En el nacimiento de un ser humano, éste puede ser niño o niña sin que sea necesario señalar cuál es considerado un éxito o un fracaso.

Solución:

La probabilidad de un éxito se expresa $P(E) = p$ y la de un fracaso $P(F) = 1 - p = q$

Si p es la probabilidad de que ocurra un suceso en un ensayo llamado **probabilidad de éxito** y $q = 1 - p$ es la probabilidad de que el suceso no ocurra en un ensayo y lo consideramos como probabilidad de fracaso, entonces la probabilidad de que el suceso se presente exactamente x veces en n ensayos se obtiene con la relación:

$$P(X) = {}_nC_x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

A esta relación se le llama **distribución binomial** porque para $x = 0, 1, 2, \ldots, n$, las expresiones ${}_nC_x$ corresponden a términos sucesivos de la relación $(q + p)^n = {}_nC_0 q^n + {}_nC_0^1 q^{n-1} p + {}_nC_2 q^{n-2} p^2 + \ldots + {}_nC_n q^n p^n$, donde ${}_nC_0$, ${}_nC_1$, ${}_nC_2$, son los coeficientes binomiales y q y p los parámetros.

Problema 3

Calcula la probabilidad de obtener exactamente dos caras en ocho lanzamientos de moneda.

Solución:

Sustituimos en $P(X) = {}_nC_x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$ con:

$n = 8$

$x = 2$

$q = p = \frac{1}{2}$, ya que la probabilidad de que caiga cara es $\frac{1}{2}$, igual que la de obtener cruz.

$$P(X) = {}_8C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{8-2} = \frac{8!}{2!(8-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{8-2} = \frac{\overset{4}{\cancel{8}}(7)(\cancel{6}!)}{\cancel{2}(1)(\cancel{6}!)} \left(\frac{1}{4}\right)\left(\frac{1}{64}\right)$$

$$= \frac{28}{256} = 0.1093$$

Problema 4

Calcula la probabilidad de obtener al menos tres caras en cinco lanzamientos de moneda.

Solución:

Sustituimos en:

$P(X) = {}_nC_x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$ con:

$n = 5$ $\quad$ $n = 5$ $\quad$ $n = 5$

$x = 3$ $\quad$ $x = 4$ $\quad$ $x = 5$

$$q = p = \frac{1}{2}$$

Expresamos la operación:

$$P(X) = {}_5C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3} + {}_5C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{5-4} + {}_5C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{5-5}$$

Calculamos por separado cada término:

$${}_5C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5!}{3!(5-3)!}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5(\overset{2}{\cancel{4}})(3!)}{3!(2)(1)}\left(\frac{1}{8}\right)\left(\frac{1}{4}\right) = \frac{10}{32}$$

$${}_5C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right) = \frac{5!}{4!(5-4)!}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right) = \frac{5(\cancel{4}!)}{\cancel{4}!(1)!}\left(\frac{1}{16}\right)\left(\frac{1}{2}\right) = \frac{5}{32}$$

$${}_5C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = \frac{\cancel{5}!}{\cancel{5}!(5-5)!}\left(\frac{1}{32}\right)(1) = \frac{1}{32}$$

$$P(X) = \frac{10}{32} + \frac{5}{32} + \frac{1}{32} = \frac{16}{32} = 0.5000$$

Problema 5

Si 20% de las piezas de televisión que fabrica una maquinaria recién ajustada son defectuosas, calcula la probabilidad de que en cinco piezas elegidas al azar se obtenga:

a) Una pieza defectuosa.

b) Ninguna defectuosa.

c) Dos piezas defectuosas como máximo.

Solución:

Sustituimos en $P(X) = {}_nC_x p^x q^{n-x} = \dfrac{n!}{x!(n-x)!} p^x q^{n-x}$ con:

$n = 5$ $\quad$ $p = 20\% = 0.2$ piezas defectuosas

$x = 1$ $\quad$ $q = 1 - p = 1 - 0.2 = 0.8$ no defectuosas

a) Una pieza defectuosa $= {}_5C_1(0.2)^1(0.8)^{5-1} = \dfrac{5!}{1!(5-1)!}(0.2)(0.8)^4$

$$= \frac{5(\cancel{4}!)}{\cancel{4}!}(0.2)(0.4096) = 0.4096$$

b) Ninguna pieza defectuosa.

Sustituimos con:

$n = 5$ $p = 0.2$ recuerda $0! = 1$

$x = 0$ $q = 0.8$

$$p(\text{cero pieza defectuosas}) \; {}_5C_0(0.2)^0(0.8)^{5-0} = \frac{5!}{0!(5-0)!}(0.3276)$$

$$= \frac{\cancel{5!}}{\cancel{5!}}(0.32768) = 0.3276$$

c) Dos piezas defectuosas como máximo.

Sustituimos con:

$n = 5$ $p = 0.2$

$x = 2$ $q = 0.8$

Primero calculamos

$$P(\text{dos piezas defectuosas}) = {}_5C_2(0.2)^2(0.8)^3 = \frac{5!}{2!(5-2)!}(0.2)^2(0.8)^3$$

$$= \frac{5(\overset{2}{\cancel{4}})(\cancel{3!})}{\cancel{2}(1)(\cancel{3!})}(0.04)(0.512) = 0.2048$$

Ahora calculamos:

P(dos piezas defectuosas como máximo) $-$ P(0 piezas defectuosas)

$+ P$ (una pieza defectuosa) $+ P$ (dos piezas defectuosas)

$= 0.32768 + 0.4096 + 0.2048$

$= 0.94208$

B. Algunas propiedades de la distribución binomial son:

La media $u = np$

Varianza $\sigma^2 = npq$

Desviación $\sigma = \sqrt{npq}$

Donde n es el número de ensayos.

En ocasiones resultan engorrosos los cálculos para la solución de algunos problemas donde intervienen probabilidades binomiales. Cuando el valor de n es muy grande y el de p es proporcionalmente menor, la distribución binomial se aproxima a la desviación de Poisson. Para reducir el tiempo que se necesita para resolverlos se han desarrollado tablas que obtienen directamente la probabilidad de r o más éxitos en n ejecuciones independientes, y cualquiera de ellas con probabilidad de éxito p.

Las probabilidades que se citan en la tabla se conocen como **probabilidades acumulativas** porque se obtuvieron por adición o acumulación de varias probabilidades separadas.

Distribución de Poisson

A. Se considera a la distribución de Poisson como una forma límite de la binomial cuando $n \Rightarrow \infty$; sin embargo, también es posible considerarla en sí misma como un proceso de Poisson.

Ambas distribuciones, junto con la normal, se aplican en la industria (en el control de calidad), en biología (para determinar el número de bacterias), en física (para calcular las partículas emitidas por una sustancia radiactiva), en las instituciones de seguros (para verificar el número de accidentes) y en otros campos más.

Las características de esta distribución son:

a) En el proceso que se estudia se identifica una unidad que puede ser de tiempo, espacio, volumen, etcétera.

b) Se contabiliza un cierto número de ocurrencias eventuales para cada unidad.

c) La variable aleatoria puede tomar una cantidad infinita pero numerable de valores $X = 0, 1, 2, ...$

Ejemplos:

- **1.** Unidad: un litro.
  - Ocurrencia eventual: presencia de bacterias de cólera.
  - Proceso con distribución de Poisson: calcular el número de bacterias por litro que hay en el agua de una delegación política.
- **2.** Unidad: 24 horas.
  - Ocurrencia eventual: robo de vehículos.
  - Proceso con distribución de Poisson: calcular el número de vehículos robados cada 24 horas en una ciudad.
- **3.** Unidad: la página de un libro.
  - Ocurrencia eventual: erratas detectadas en un libro.
  - Proceso con distribución de Poisson: las erratas que aparecen en cada página de un libro recién publicado.
- **4.** Unidad: Tinacos de agua con capacidad de 1 000 litros.
  - Ocurrencia eventual: consumo de agua.
  - Proceso con distribución de Poisson: la cantidad de tinacos de agua potable consumidos por las escuelas primarias de una ciudad.

B. Un problema que satisface las características anteriores se resuelve con la distribución de probabilidades de Poisson y con la relación $P(X) = \dfrac{\lambda^x e^{-\lambda}}{x!}$, donde:

El número irracional $e = 2.71828...$

La letra landa (λ), del alfabeto griego, es el parámetro que determina el valor de esta distribución.

C. Algunas propiedades de la distribución de Poisson son:

La media $u = \lambda$

Varianza $\sigma^2 = \lambda$

Desviación $\sigma = \sqrt{\lambda}$

D. Uso de la distribución de Poisson.

Cuando aplicamos la relación de la distribución binomial:

$$P(X) = {}_nC_x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

como en el caso siguiente:

Problema 6

Si la probabilidad de que una persona sufra una reacción negativa al ingerir un determinado antibiótico es de 0.001, calcula la probabilidad de que un total de 3 000 pacientes sufran una reacción negativa:

a) Exactamente tres personas sufran reacción negativa.

b) Más de dos personas presenten reacción negativa.

Solución:

1. Si aplicamos la distribución de Poisson tenemos:

$$P(X \text{ personas sufran reacción negativa}) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Con $\lambda = np = 3\,000(0.001) = 3;\ x = 3$

Entonces:

a) Exactamente tres personas sufran reacción negativa.

P (tres personas sufran reacción negativa) =

$$\frac{3^3 e^{-3}}{3!} = \frac{27}{3(2)(1)e^3} = \frac{27}{6(20.083)}$$

$$= 0.2240$$

b) Más de dos personas presenten reacción negativa.

P (más de dos personas sufran reacción negativa)

P (ninguna persona sufra reacción negativa) $= \frac{2^0 e^{-3}}{0!} = \frac{1}{e^3}$

P (una persona sufra reacción negativa) $= \frac{2^1 e^{-3}}{1!} = \frac{2}{e^3}$

P (dos personas sufran reacción negativa) $= \frac{2^2 e^{-3}}{2!} = \frac{4}{2(1)e^3} = \frac{2}{e^3}$

P (más de 2 personas sufran reacción) = 1 − (0 o 1 o 2 sufran reacción negativa)

$$= 1 - \left(\frac{1}{e^3} + \frac{2}{e^3} + \frac{2}{e^3} \right) = 1 - \frac{5}{e^3} = 1 - \frac{5}{20.08}$$

$$= 0.751$$

Si aplicáramos la distribución binomial para resolver el problema, tendríamos:

a) Exactamente tres personas sufran reacción negativa.

$${}_{3000}C_3(0.001)^3(0.999)^{1997}$$

b) Más de dos personas presenten reacción negativa.

$$1 - \{{}_{3000}C_0(0.001)^0(0.999)^{2000} + {}_{3000}C_1(0.001)^1(0.999)^{1999}$$
$$+ {}_{3000}C_2(0.001)^2(0.999)^{1998}\}$$

Este resultado es muy difícil y engorroso de obtener.

Conclusión:

Como en el problema anterior, cuando el valor de n es grande y la probabilidad p de ocurrencia del suceso está cerca de cero, de modo que el resultado de $q = 1 - p$ está cerca de 1, la solución resulta difícil y engorrosa.

En general, se aplica la distribución de Poisson cuando el número de repeticiones del experimento es al menos de 50($n \geq 50$) y np es menor que 5. En estos casos, la distribución binomial se aproxima a la distribución de Poisson y cuando ésta se aplica, generalmente se conocen los valores de n y p.

Si aplicamos la distribución de Poisson en sí misma, no se conoce el valor de p y es necesario que calculemos el de λ.

E. Tabla de valores de $e^{-\lambda}$, su aplicación.

Al usar la distribución de Poisson los cálculos se facilitan si los valores de $e^{-\lambda}$ se obtienen directamente de una tabla que incluye, en una primera parte, los valores donde λ es de ($0 < \lambda < 1$) y en una segunda, los de $\lambda = 1, 2, 3, \ldots 10$.

Primera parte

Tabla de valores de $e^{-\lambda}$ con ($0 < \lambda < 1$)

x	0	1	2	3	4
λ					
0.0	0.0000	0.9900	0.9802	0.9704	0.9608
0.1	0.9048	0.8958	0.8869	0.8781	0.8694
0.2	0.8187	0.8106	0.8025	0.7945	0.7866
0.3	0.7408	0.7334	0.7261	0.7189	0.7118
0.4	0.6703	0.6636	0.6570	0.6505	0.6440
0.5	0.6065	0.6005	0.5945	0.5886	0.5827
0.6	0.5488	0.5434	0.5379	0.5326	0.5273
0.7	0.4966	0.4916	0.4868	04819	0.4771
0.8	0.4493	0.4449	0.4404	0.4360	0.4317
0.9	0.4066	0.4025	0.3985	0.3946	0.3906

x	5	6	7	8	9
λ					
0.0	0.9512	0.9418	0.9324	0.9231	0.9139
0.1	0.8607	0.8521	0.8437	0.8353	0.8270
0.2	0.7788	0.7711	0.7634	0.7558	0.7483
0.3	0.7047	0.6977	0.6907	0.6839	0.6771
0.4	0.6376	0.6313	0.6250	0.6188	0.6126
0.5	0.5770	0.5712	0.5655	0.5599	0.5543
0.6	0.5220	0.5169	0.5117	0.5066	0.5016
0.7	0.4724	0.4677	0.4630	0.4584	0.4538
0.8	0.4274	0.4232	0.4190	0.4148	0.4107
0.9	0.3867	0.3829	0.3791	0.3753	0.3716

Segunda parte

Para ($\lambda = 1, 2, 3, ..., 10$)

λ	1	2	3	4	5
$e^{-\lambda}$	0.36788	0.13534	0.04979	0.01832	0.006738

λ	6	7	8	9	10
$e^{-\lambda}$	0.002479	0.000912	0.000335	0.000123	0.000045

Para ($\lambda = 1, 2, 3, ..., 10$)

Para aplicar la segunda parte de la tabla y para calcular $e^{-\lambda}$ usamos las leyes de los exponentes.

Ejemplo:

1. Determina el valor de $e^{-5.35}$

Solución:

$$e^{-5.35} = (e^{-5.00})(e^{-0.35}) = (0.006738)(0.7047)$$

$$= 0.0047482$$

En la tabla de $\lambda = 1, 2, 3, ..., 10$ localizamos el valor de -5 y en la de $(0 < \lambda < 1)$ el de -0.35. Los resultados los multiplicamos.

Problema 7

Si una distribución de Poisson es de $P(X) = \dfrac{(0.56)^x e^{-0.56}}{x!}$, calcula:

a) $P(0)$

b) $P(1)$

c) $P(3)$

d) $P(4)$

Solución:

a) $P(0) = \dfrac{(0.56)^0 e^{-0.56}}{0!} = \dfrac{(1)e^{-0.56}}{1} = e^{-0.56} = 0.5712$

Obtuvimos el resultado consultando la tabla en la primera parte.

b) $P(1) = \dfrac{(0.56)^1 e^{-0.56}}{1!} = 0\ 56e^{-0.56} = 0\ 56(0.5712) = 0.3198$

c) $P(3) = \dfrac{(0.56)^3 e^{-0.56}}{3!} = \dfrac{0.1756}{3(2)(1)} e^{-0.56} = 0.0292(0.5712) = 0.0166$

d) $P(4) = \dfrac{(0.56)^4 e^{-0.56}}{4!} = \dfrac{0.0983}{4(3)(2)(1)} e^{-0.56} = 0.0040(0.5712) = 0.0022$

(se tomaron cuatro cifras decimales)

Problema 8

Calcula el valor $e^{-5.28}$ utilizando la tabla para valores $\lambda = 1, 2, 3, ..., 10$.

Solución:

Para obtener el resultado aplicamos las leyes de los exponentes:

$e^{-5.28} = (e^{-5.00})(e^{-0.28}) = (0.006738)(0.7558) = 0.0050$

(se tomaron 4 cifras decimales)

Ejercicios de repaso

1. Determina el valor de las expresiones siguientes:

a) 4!

b) $\dfrac{6!}{2!(5!)}$

c) ${}_5C_2$

d) ${}_3C_3$

e) ${}_5C_0$

Sol. 24, 3, 10, 1, 1

2. Calcula la probabilidad de que al lanzar una moneda al aire se obtenga:

a) 3 caras

b) 2 caras

Sol. 0.25, 0.375

3. En una fábrica de ropa, 10% de las prendas producidas resultan con algún defecto. Calcula la probabilidad de que en un lote de 9 prendas elegidas al azar salgan exactamente 2 defectuosas.

Sol. 0.1721

4. La probabilidad de que un técnico en computación perciba un sueldo mayor a 10 000 pesos mensuales es de 0.001. Calcula la probabilidad de que en un total de 2 000 técnicos cuatro perciban exactamente este sueldo. Usa la relación de la distribución de Poisson.

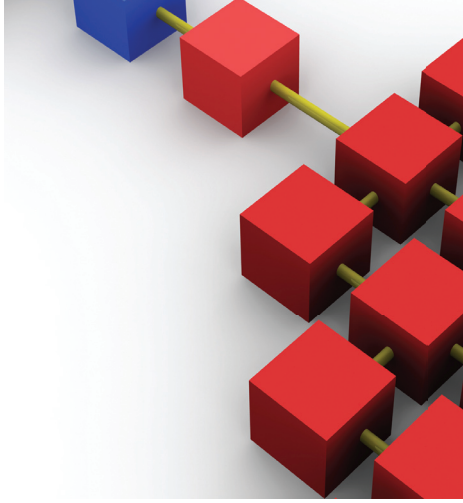
Sol. 0.0902

5. Si una distribución de Poisson es de $P(X)=\dfrac{(0.35)^x e^{-0.35}}{x!}$ calcula:

a) $P(0)$

b) $P(3)$

Sol. **a)** 0.7047
b) 0.0050



Capítulo 13

Distribución de probabilidades continuas

Variable normalizada

Distribución normal

Conceptos clave

Variable normalizada o calificación estándar
Distribución normal o curva normal

Variable normalizada. Calificación estándar *Z*

Definimos la media aritmética como la suma de los valores de cierto número de cantidades dividida entre su número. Se expresa así:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{N}$$

La media aritmética es la única medida de tendencia central que puede intervenir en operaciones algebraicas.

A la desviación estándar, o desviación típica, la definimos como la raíz cuadrada de la varianza. Se expresa así:

$$s = \sqrt{s^2}$$

La desviación estándar es la más importante de todas las medidas de dispersión porque incluye más o menos 68% de los términos de una distribución normal. Por sus propiedades, se utiliza con facilidad en el análisis estadístico.

A continuación analizaremos la **variable normalizada** o **calificación estándar**, que es el resultado de dividir la media aritmética entre la desviación estándar. También se le denomina variable abreviada calificación *Z*.

$$\text{Calificación } Z = \frac{\text{Desviación respecto a la media aritmética}}{\text{Desviación estándar}}$$

Se expresa así:

$$Z = \frac{X - \overline{X}}{s}$$

Sin embargo, algunos autores la citan así:

$$Z = \frac{x - \mu}{\sigma}$$

donde la media aritmética $\overline{X}$ se representa con la letra griega μ (mu) y la desviación estándar con la letra griega σ (ro).

La calificación Z es adimensional y por ello su aplicación es útil para comparar datos individuales de distribuciones que tienen distintas unidades de medida, así como diferentes medidas y desviaciones estándar.

El resultado expresa la desviación de la calificación en unidades de desviación estándar.

Problema 1

Calcula el valor de la calificación Z de una serie de números cuya media aritmética $\bar{X} = 110$ y desviación estándar $s = 6$, para una observación de puntuación 134.

Solución:

Desviación $X - \bar{X} = 134 - 110 = 24$

Calificación $Z = \dfrac{24}{6} = 4$

Se interpreta así: la calificación 134 está 4 desviaciones estándar por encima de la media aritmética.

Para una calificación de $110 - 6 = 104$: $Z = \dfrac{104 - 110}{6} = \dfrac{-6}{6} = -1$ y decimos:

La calificación 104 está 1 desviación estándar abajo de la media.

Propiedades de la calificación estándar

a) La suma de las calificaciones Z es 0.

Como $\sum (X - \bar{X}) = 0$, entonces $\dfrac{\sum (X - \bar{X})}{s} = \dfrac{0}{s} = 0$

b) La media de las calificaciones $Z = 0$.

$\bar{Z} = 0$ como corolario de la propiedad **a)**.

c) La suma de los cuadrados de las calificaciones Z es igual a N.

$\sum Z^2 = N$

Ya que:

$$\sum Z^2 = \frac{\sum (X - \bar{X})^2}{s^2} = \frac{1}{s^2} \sum (X - \bar{X})^2$$

como $s^2 = \dfrac{\sum (X - \bar{X})^2}{s^2}$, sustituimos:

$$\sum Z^2 = \frac{N}{\cancel{\sum (X - \bar{X})^2}} \cancel{\sum (X - \bar{X})^2}$$

$\sum Z^2 = N$

d) La desviación estándar y la varianza de las calificaciones Z es 1.

$$\sigma_z^2 = \frac{\sum (Z - Z)^2}{N}$$

como $\overline{Z} = 0$ y $\sum Z^2 = N$, sustituyendo queda:

$$\sigma_z^2 = \frac{N}{N} = 1$$

Nota: $\sigma = s$ desviación estándar.

Problema 2

Después de realizar una campaña de promoción de la salud en una escuela, se presentaron los siguientes datos sobre la población de alumnos:

- Para la estatura media $\overline{X} = 164.2$. La desviación estándar $s = 8.34$.
- Para los pesos, la media $\overline{X} = 60.2$. La desviación estándar $s = 12.4$.

La estatura de Juan es de 169 cm y pesa 65 kg.

Calcula la calificación Z para cada medida. ¿Cuál de las medidas tiene la mayor desviación y cómo consideras la estatura y el peso de Juan?

Solución:

Calificación Z:

$$\text{Estatura } Z = \frac{169 - 164.2}{8.34} = 0.57$$

$$\text{Peso } Z = \frac{65 - 60.2}{12.4} = 0.38$$

La mayor desviación es la que corresponde a la estatura.

La estatura y el peso de Juan están correctos.

Problema 3

Calcula la calificación Z de cada puntuación de la serie 4, 5, 6, 10, 8, 7, 2, 1

Solución:

Calculamos la media aritmética $\overline{X} = \mu$

$$\overline{X} = \frac{\sum X}{N} = \frac{4+5+6+10+8+7+2+1}{8} = \frac{43}{8} = 5.375$$

Desviación estándar $\sigma = s$

$$s = \sqrt{\frac{\sum X^2}{N} - \overline{X}^2} = \sqrt{\frac{4^2+5^2+6^2+10^2+8^2+7^2+2^2+1^2}{8} - (5.375)^2}$$

$$= \sqrt{\frac{16+25+36+100+64+49+4+1}{8} - 28.89} = \sqrt{\frac{295}{8} - 28.89}$$

$$= \sqrt{36.875 - 28.89} = \sqrt{7.985} = 2.82 \text{ (se tomaron dos cifras decimales)}$$

Obtenemos la calificación Z de cada puntuación:

Puntuación	$\frac{X-\overline{X}}{s}=Z$
4	$(4-5.375)\div 2.82=-0.48$
5	$(5-5.375)\div 2{,}82=-0.13$
6	$(6-5.375)\div 2.82=0.22$
10	$(10-5.375)\div 2.82=1.64$
8	$(8-5.375)\div 2.82=0.93$
7	$(7-5.375)\div 2.82=0.57$
2	$(2-5.375)\div 2.82=-1.19$
1	$(1-5.375)\div 2.82=-1.55$

La aplicación principal de la calificación Z se hace en el estudio de la distribución normal.

Distribución normal

Uno de los ejemplos más importantes de una probabilidad continua es la **distribución normal** o **curva normal**.

En el campo de la estadística, la distribución normal es la más importante de las distribuciones de frecuencias, ya que la mayoría de los procedimientos estadísticos se basan en ella.

A pesar de que el estudio formal de la curva normal y la tabla de áreas corresponde a un nivel superior, a continuación daremos una explicación conceptual para así poder calcular la probabilidad de un suceso.

Debido a la intervención del matemático Karl Gauss (1777–1855) en el estudio de la distribución normal, algunos autores la denominan **distribución gaussiana**.

La distribución normal se da con la relación:

$$Y=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}Z^2}$$

Donde:

σ = desviación estándar

$\pi = 3.1416$

$e = 2.71828...$

Z = Variable normalizada (calificación estándar Z)

A la relación citada para obtener la distribución normal se le conoce como forma tipificada y se dice que Z se distribuye normalmente con la media cero y varianza uno.

El área total limitada por la curva y el eje de las x es uno; de ahí que el área bajo la curva entre dos ordenadas $x = a$ y $x = b$, donde $a < b$ representa la probabilidad de que X se encuentre entre a y b, se expresa $P(a < X < b)$.

Algunas propiedades de la distribución normal dada por la relación antes citada son:

La media es μ

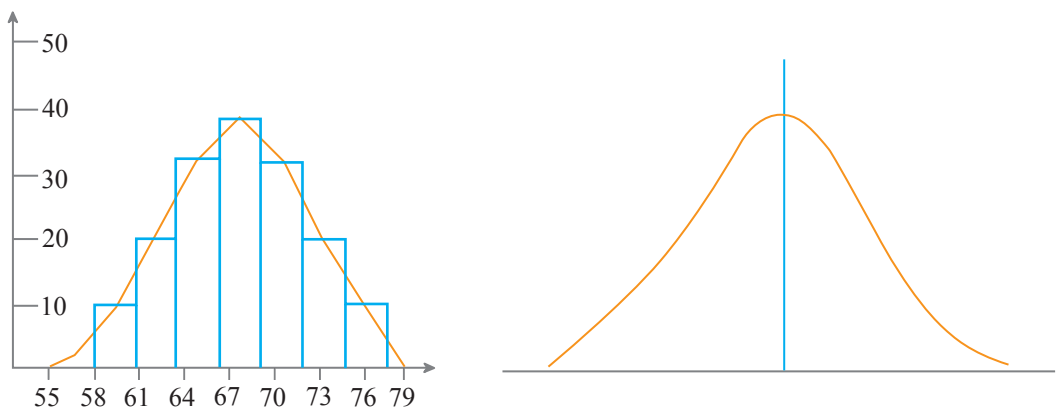
La varianza es σ^2

La desviación típica es σ

Desviación media $\sigma\sqrt{\frac{2}{\pi}} = 0.7979\sigma$

Si en el histograma de distribución de frecuencias:

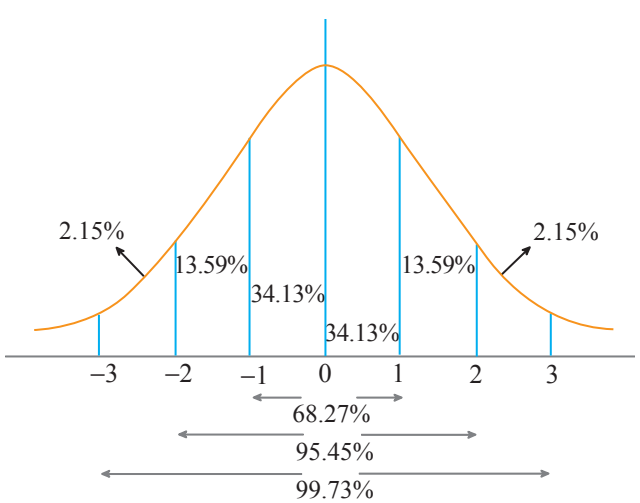
Se aumenta la cantidad de observaciones, entonces los intervalos de clase se hacen más angostos y el polígono de frecuencias se transforma en una curva suave.



Propiedades de la curva normal

1. Es simétrica y tiene forma de campana.
2. La media aritmética está a la mitad y divide el área en dos mitades.
3. Teóricamente, la curva se extiende en ambas direcciones y tiende gradualmente a unirse con la recta horizontal hasta el infinito sin tocarla nunca, es asíntota.

La curva normal se acepta como modelo ideal de una situación real. Su uso y el de la tabla de áreas normales permiten obtener la probabilidad de un suceso.



Observando la figura aceptamos:

A. El punto medio de la curva normal es la media aritmética

$$\bar{X} = \mu \text{ y es igual a cero}$$

$$\mu = 0$$

B. La desviación estándar $s = \sigma$ y es igual a 1.

$$\sigma = 1$$

El área total limitada por la curva y el eje de las abscisas es igual a 1 (desviación estándar σ) y equivale al 100% de los casos; de tal manera que la proporción de área bajo la curva limitada por dos ordenadas (perpendiculares) levantadas en puntos del eje de las abscisas, expresa el porcentaje de casos comprendidos entre las calificaciones Z correspondientes a los dos puntos en que se trazaron.

Entre la media $\mu = 0$ y el punto 1 de desviación estándar se encuentra el 34.13% del total de casos y el área bajo la curva es igual a 0.3413.

Entre la media $\mu = 0$ y el punto -1 de desviación estándar está otro 34.13% de todos los casos y el área bajo la curva es igual a 0.3413.

De donde, entre $Z = -1$ y $Z = +1$ se encuentra el 68.27% del total de todos los casos, y el área bajo la curva es de 0.6827.

Entre la media $\mu = 0$ y el punto 2 de desviación estándar se encuentra 34.13 + 13.59 = 47.72% de todos los casos y el área bajo la curva = 0.3413 + 0.1359 = 0.4772.

Entre la media $\mu = 0$ y el punto -2 de desviación estándar, se encuentran otros 34.13 + 13.59 = 47.72% de todos los casos, y el área bajo la curva es igual a 0.3413 + 0.1359 = 0.4772.

Por lo tanto, entre $Z = -2$ y $Z = +2$ se encuentran el 95.45% de todos los casos, y el área bajo la curva = 0.4772 + 0.4772 − 0.9544.

Tabla de áreas bajo la curva normal. Cómo usarla

La tabla de áreas bajo la curva normal desde la media 0 a Z que se cita al final de este apartado fue elaborada por varios matemáticos y se incluye en los libros de estadística de enseñanza superior. Esta tabla se aplica así.

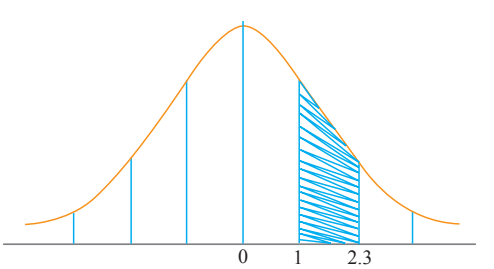
Área bajo la curva

Problema 4

Calcula el área bajo la curva normal entre $Z = 1$ y $Z = 2.3$

Solución:

Trazamos la curva:



Razonando:

Desde $Z = 0$ hasta $Z = 2.3$ el área bajo la curva $= 0.4893$

Desde $Z = 0$ hasta $Z = 1$ el área bajo la curva $= 0.3413$

Diferencia entre las áreas $= 0.1480$

El área es igual a 0.1480 y representa la probabilidad de que Z esté comprendida entre 1 y 2.3; se expresa $P(1 \leq Z \leq 2.3)$.

Para obtener este resultado seguimos este procedimiento:

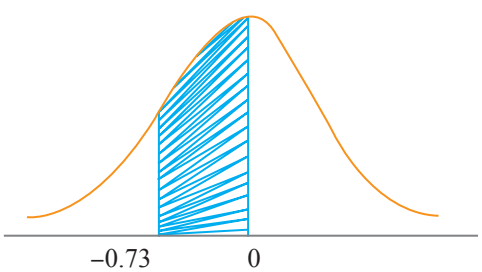
- En la tabla citada localizamos el valor 2.3 y a su derecha está su valor 0.4893.
- Del mismo modo obtuvimos el valor que corresponde, 0.3413.
- Buscamos los valores en la tabla para resolver los problemas siguientes.

Problema 5

Calcula el área bajo la curva normal entre $Z = 0.73$ y $Z = 0$.

Solución:

Trazamos la curva:



Razonando:

El área entre −0.73 y 0 es 0.2673

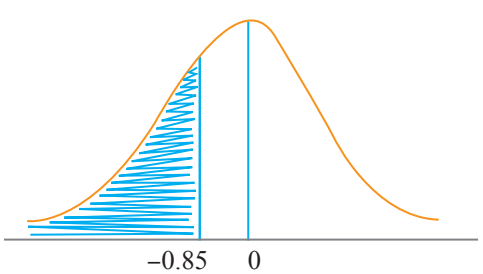
El área es igual a 0.2673 y representa la probabilidad de que Z esté comprendida entre −0.73 y 0. Se expresa $P(-0.73 \leq Z \leq 0)$.

Problema 6

Calcula el área bajo la curva normal a la izquierda de $Z = -0.85$.

Solución:

Trazamos la curva:



Razonando:

Área a la izquierda de $Z = 0$ es	0.5000
Área desde $Z = 0$ hasta $Z = -0.85$ es	0.3023
Diferencia entre las áreas	0.1977

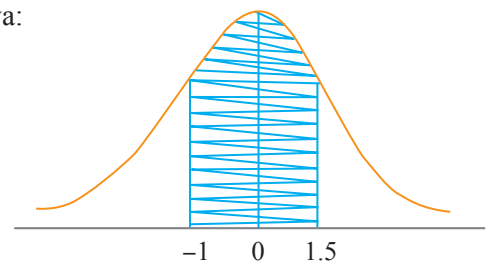
El área es igual a 0.1977 y representa la probabilidad de que Z esté comprendida a la izquierda de $Z = -0.85$. Se expresa $P(Z \leq -0.85)$.

Problema 7

Calcula el área bajo la curva normal entre $Z = -1$ y $Z = 1.5$

Solución:

Trazamos la curva:



Razonando:

Desde $Z = -1$ hasta $Z = 0$ el área bajo la curva es	0.3413
Desde $Z = 0$ hasta $Z = 1.5$ el área bajo la curva es	0.4332
Suman las áreas	0.7745

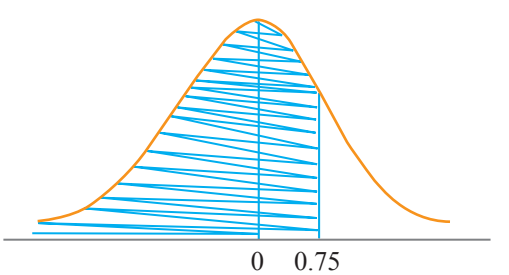
El área es igual a 0.7745 y representa la probabilidad de que Z esté comprendida entre -1 y 1.5. Se expresa $P(-1 \leq Z \leq 1.5)$.

Problema 8

Calcula el área bajo la curva normal a la izquierda de $Z = 0.75$.

Solución:

Trazamos la curva:



Razonando:

Desde $Z = 0$ hasta $Z = 0.75$ el área bajo la curva es	0.2734
Área a la izquierda de $Z = 0$ es	0.5000
Suman las áreas	0.7734

El área es igual a 0.7734 y representa la probabilidad de que Z esté comprendida a la izquierda de $Z = 0.75$. Se expresa $P(Z \leq 0.75)$.

En los problemas anteriores calculamos las áreas bajo la curva normal; ahora, si conocemos el área vamos a obtener el valor de Z.

Cálculo del valor o valores de Z

Problema 9

Obtener el valor de Z si el área que corresponde a una curva normal entre 0 y Z es de 0.2540.

Solución:

Buscamos en la tabla de áreas bajo la curva normal el valor más próximo a 0.2540, que corresponde a $Z = 0.69$, pero como Z puede estar a la izquierda o a la derecha de 0 para la misma área, la respuesta es $Z = \pm 0.69$ con una aproximación de centésimos.

Si deseamos un valor aproximado a diezmilésimos recurrimos a la interpolación lineal entre los dos valores más próximos a 0.2540 en la forma siguiente:

a	0.2549	corresponde	0.69
a	0.2517	corresponde	0.68
	0.0032		0.01

Además, la diferencia entre 0.2549 y 0.2517 es 0.0032. Entonces:

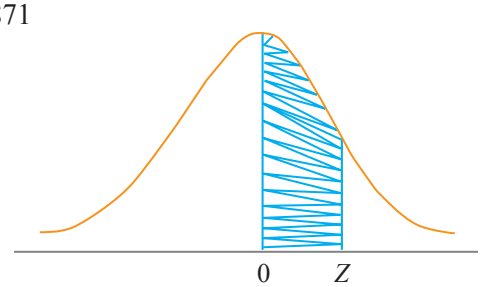
$$x : 01 : : 0.0023 : 0.0032$$

$$x(0.0032) = 0.01(0.0023)$$

$$x = \frac{0.01(0.0023)}{0.0032} = 0.0071$$

Por lo tanto:

$$Z = 0.68 + 0.0071 = 0.6871$$



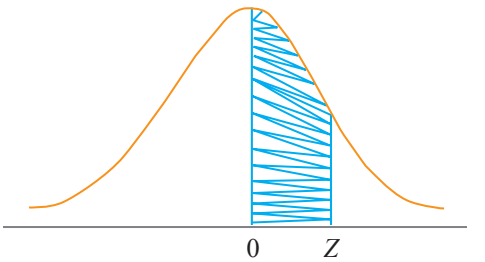
± 0.6871

Problema 10

Determina el valor de Z si el área que le corresponde a una curva normal entre 0 y Z es de 0.4545.

Solución:

En la tabla de áreas bajo curva normal, el número 0.4545 se localiza a la derecha del 1.69. Entonces:



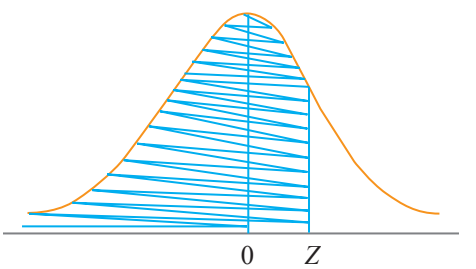
$Z = 1.69$

Problema 11

Determina el valor de Z si el área que corresponde a una curva normal, cuya área está a la izquierda de Z, es de 0.8729.

Solución:

Como Z es positivo y el área es mayor que 0.5000, el área entre 0 y Z = 0.8729 – 0.5000 = 0.3729.



En la tabla de área bajo la curva normal, el número 0.3729 se localiza a la derecha del 1.14.

Problema 12

El coeficiente intelectual de los aspirantes aprobados para ingresar a la Escuela Médico Militar tiene una media aritmética de $\mu = 100$ una desviación estándar de $s = \sigma = 10$.

Calcula cuál es la proporción de reclutas que tienen un coeficiente intelectual entre 100 y 107 en términos de la calificación estándar Z.

Solución:

$$Z = \frac{X - \bar{X}}{s} \text{ o } Z = \frac{x - \mu}{\sigma} = \frac{107 - 100}{10} = \frac{7}{10} = 0.70$$

La desviación típica es de 0.70. A continuación buscamos en la tabla de áreas normales y obtenemos 0.2580, de donde la proporción que nos interesa la obtenemos así:

0.5000 – 0.2580 = 0.2420

Interpretamos:

0.2420 = 24.20%

El 24.20% de los alumnos tienen un coeficiente intelectual entre 100 y 107.

Para expresar el resultado anterior, en términos de enunciados de probabilidad, expresamos:

$$Z = \frac{x - \mu}{\sigma} \text{ o } Z = \frac{X - \bar{X}}{s}$$

$$Z = \frac{x - \mu}{\sigma} = \frac{107 - 100}{10} = \frac{7}{10} = 0.70 \text{ , que podemos expresar:}$$

Probabilidad $P(x \geq 107)$

$$= P\left(\frac{x - 100}{10} \geq \frac{107 - 100}{10} \right)$$

$$= P(Z \geq 0.70)$$

$$= 0.2580$$

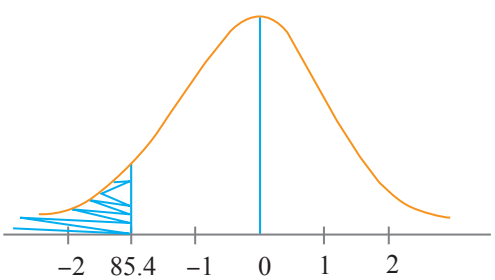
Problema 13

Considerando el tipo de alumnos citados en el problema anterior, calcula qué proporción de ellos tienen un coeficiente intelectual menor a 85.4

Solución:

$$Z = \frac{x - \mu}{\sigma} = \frac{85.4 - 100}{10} = -\frac{14.6}{10} = -1.46$$

Para comprender mejor el resultado trazamos una curva donde lo expresamos.



Ahora citamos el resultado en términos de enunciados de probabilidad para expresar la relación entre la probabilidad y la curva normal de áreas.

Probabilidad $P(x \leq 85.4)$

$$= P\left(\frac{x - 100}{10} \geq \frac{85.4 - 100}{10} \right)$$

$$= P(Z \leq 1.46)$$

$$= 0.0721$$

El valor de la probabilidad 0.0721 la obtuvimos así:

- Señalamos que el área total limitada por la curva es igual al equivalente al 100% de los casos, la perpendicular que está en 0 divide en dos partes el área:

  $0.5000 + 0.5000 = 1.000 = 1$

- Obtuvimos en la tabla el valor de $Z = 1.46$, que es 0.4279. Por lo tanto:

  $0.5000 - 0.4279 = 0.0721$

Problema 14

Con los datos de los alumnos citados en el problema 12, calcula la proporción de alumnos con un coeficiente intelectual mayor que 108, donde $\mu = 100$ y $\sigma = 10$

Solución:

$$Z = \frac{x - \mu}{\sigma} = \frac{108 - 100}{10} = \frac{8}{10} = 0.8$$

En la tabla normal de áreas se tiene que la proporción de Z es igual a 0.80 y su área es de 0.2881, de donde:

$0.5000 - 0.2881 = 0.2119$

Cálculo del rango percentil

Si conocemos la media aritmética y la desviación estándar, y necesitamos calcular el rango percentil de una puntuación dada con la curva normal, el problema se resuelve obteniendo el área bajo la curva.

Problema 15

En un examen departamental de una escuela de enseñanza media superior, la media aritmética de las calificaciones es de 68 y la desviación estándar de 17. Calcula:

a) El rango percentil para una calificación de 77.

b) El rango percentil de una calificación de 40.

c) El tanto por ciento de los alumnos que obtuvieron calificaciones entre 76 y 81.

Solución:

a) Pasamos la calificación 77 a calificación Z:

$$Z = \frac{77 - 68}{17} = \frac{9}{17} = 0.5294 = 0.53$$

De la tabla determinamos el área entre $Z = 0$ y $Z = 0.53$, que es de 0.2019. Dado que el valor de 0.53 está a la derecha de 0, debemos sumar 0.5, de donde:

0.5000

0.2019

0.7019

Rango percentil = 70.19%

b) Pasamos la calificación 40 a calificación Z:

$$Z = \frac{40 - 68}{17} = \frac{-28}{17} = -1.64 \text{ (se tomaron dos cifras decimales)}$$

De la tabla determinamos el área entre $Z = 0$ y 1.64, que es de 0.4495. Dado que Z es negativo, se resta el valor del área a 0.5000:

0.5000

0.4495

0.0505

Rango percentil = 5.05%

c) Calculamos las calificaciones Z.

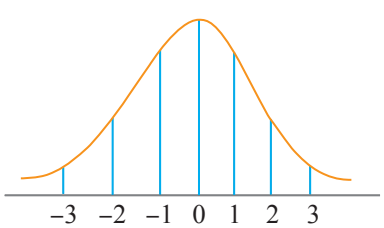
$$Z = \frac{76 - 68}{17} = \frac{8}{17} = 0.47 \text{ (se tomaron dos cifras decimales)}$$

$$Z = \frac{81 - 68}{17} = \frac{13}{17} = 0.76 \text{ (se tomaron dos cifras decimales)}$$

Obtenemos las áreas:

Desde $Z = 0$ hasta $Z = 0.76$, es de	0.2764
Desde $Z = 0$ hasta $Z = 0.47$, es de	0.1882
La diferencia entre las áreas	0.0882

8.82% de los alumnos obtuvieron calificaciones entre 76 y 81.



Áreas bajo la curva normal desde la media 0 a *Z*

Z	Área	*Z*	Área	*Z*	Área	*Z*	Área
0.00	0.0000	0.26	0.1026	0.52	0.1985	0.78	0.2823
0.01	0.0040	0.27	0.1064	0.53	0.2019	0.79	0.2852
0.02	0.0080	0.28	0.1103	0.54	0.2054	0.80	0.2881
0.03	0.0120	0.29	0.1141	0.55	0.2088	0.81	0.2910
0.04	0.0160	0.30	0.1179	0.56	0.2123	0.82	0.2939
0.05	0.0199	0.31	0.1217	0.57	0.2157	0.83	0.2967
0.06	0.0239	0.32	0.1255	0.58	0.2190	0.84	0.2995
0.07	0.0279	0.33	0.1293	0.59	0.2224	0.85	0.3023
0.08	0.0319	0.34	0.1331	0.60	0.2257	0.80	0.3051
0.09	0.0359	0.35	0.1368	0.61	0.2291	0.87	0.3078
0.10	0.0398	0.36	0.1406	0.62	0.2324	0.88	0.3106
0.11	0.0438	0.37	0.1443	0.63	0.2357	0.89	0.3133
0.12	0.0478	0.38	0.1480	0.64	0.2389	0.90	0.3159
0.13	0.0517	0.39	0.1517	0.65	0.2422	0.91	0.3186
0.14	0.0557	0.40	0.1554	0.66	0.2454	0.92	0.3212
0.15	0.0596	0.41	0.1591	0.67	0.2486	0.93	0.3238
0.16	0.0636	0.42	0.1628	0.68	0.2517	0.94	0.3264
0.17	0.0675	0.43	0.1664	0.69	0.2549	0.95	0.3289
0.18	0.0714	0.44	0.1700	0.70	0.2580	0.96	0.3315
0.19	0.0753	0.45	0.1736	0.71	0.2611	0.97	0.3340
0.20	0.0793	0.46	0.1772	0.72	0.2642	0.98	0.3365
0.21	0.0832	0.47	0.1882	0.73	0.2673	0.99	0.3389
0.22	0.0871	0.48	0.1844	0.74	0.2704	1.00	0.3413
0.23	0.0910	0.49	0.1879	0.75	0.2734	1.01	0.3438
0.24	0.0948	0.50	0.1915	0.76	0.2764	1.02	0.3461
0.25	0.0987	0.51	0.1950	0.77	0.2794	1.03	0.3485
2.08	0.4812	2.26	0.4881	2.44	0.4927	2.62	0.4956
2.09	0.4817	2.27	0.4884	2.45	0.4929	2.63	0.4957
2.10	0.4821	2.28	0.4887	2.46	0.4931	2.64	0.4959
2.11	0.4826	2.29	0.4890	2.47	0.4932	2.65	0.4960
2.12	0.4830	2.30	0.4893	2.48	0.4934	2.66	0.4961
2.13	0.4834	2.31	0.4896	2.49	0.4936	2.67	0.4962
2.14	0.4838	2.32	0.4898	2.50	0.4938	2.68	0.4963
2.15	0.4842	2.33	0.4901	2.51	0.4940	2.69	0.4964
2.16	0.4846	2.34	0.4904	2.52	0.4941	2.70	0.4965
2.17	0.4850	2.35	0.4906	2,53	0.4943	2.71	0.4966
2.18	0.4854	2.36	0.4909	2.54	0.4945	2.72	0.4967
2.19	0.4857	2.37	0.4911	2.55	0.4946	2.73	0.4968
2.20	0.4861	2.38	0.4913	2.56	0.4948	2.74	0.4969
2.21	0.4864	2.39	0.4916	2.57	0.4949	2.75	0.4970
2.22	0.4868	2.40	0.4918	2.58	0.4951	2.76	0.4971
2.23	0.4871	2.41	0.4920	2.59	0.4952	2.77	0.4972
2.24	0.4875	2.42	0.4922	2.60	0.4953	2.78	0.4973
2.25	0.4878	2.43	0.4925	2.61	0.4955	2.79	0.4974

Z	Área	*Z*	Área	*Z*	Área	*Z*	Área
1.04	0.3508	1.30	0.4032	1.56	0.4406	1.82	0.4656
1.05	0.3531	1.31	0.4049	1.57	0.4419	1.83	0.4664
1.06	0.3554	1.32	0.4066	1.58	0.4429	1.84	0.4671
1.07	0.3577	1.33	0.4082	1.59	0.4441	1.85	0.4678
1.08	0.3599	1.34	0.4099	1.60	0.4252	1.86	0.4686
1.09	0.3621	1.35	0.4115	1.61	0.4463	1.87	0.4693
1.10	0.3643	1.36	0.4131	1.62	0.4474	1.88	0.4699
1.11	0.3665	1.37	0.4147	1.63	0.4484	1.89	0.4706
1.12	0.3686	1.38	0.4162	1.64	0.4495	1.90	0.4713
1.13	0.3708	1.39	0.4177	1.65	0.4505	1.91	0.4719
1.14	0.3729	1.40	0.4192	1.66	0.4515	1.92	0.4726
1.15	0.3749	1.41	0.4207	1.67	0.4525	1.93	0.4732
1.16	0.3770	1.42	0.4222	1.68	0.4535	1.94	0.4738
1.17	0.3790	1.43	0.4236	1.69	0.4545	1.95	0.4744
1.18	0.3810	1.44	0.4251	1.70	0.4554	1.96	0.4750
1.19	0.3830	1.45	0.4265	1.71	0.4564	1.97	0.4756
1.20	0.3849	1.46	0.4279	1.72	0.4573	1.98	0.4761
1.21	0.3869	1.47	0.4292	1.73	0.4582	1.99	0.4767
1.22	0.3888	1.48	0.4306	1.74	0.4591	2.01	0.4772
1.23	0.3907	1.49	0.4319	1.75	0.4599	2.02	0.4778
1.24	0.3925	1.50	0.4332	1.76	0.4608	2.03	0.4783
1.25	0.3944	1.51	0.4345	1.77	0.4616	2.04	0.4788
1.26	0.3962	1.52	0.4357	1.78	0.4625	2.05	0.4793
1.27	0.3980	1.53	0.4370	1.79	0.4633	2.06	0.4798
1.28	0.3997	1.54	0.4382	1.80	0.4641	2.07	0.4803
1.29	0.4015	1.55	0.4394	1.81	0.4649	3.32	0.4808
2.80	0.4974	2.98	0.4986	3.15	0.4992	3.33	0.4995
2.81	0.4975	2.99	0.4986	3.16	0.4992	3.34	0.4996
2.82	0.4976	3.00	0.4987	3.17	0.4992	3.35	0.4996
2.83	0.4977	3.01	0.4987	3.18	0.4993	3.36	0.4996
2.84	0.4977	3.02	0.4987	3.19	0.4993	3.37	0.4996
2.85	0.4978	3.03	0.4988	3.20	0.4993	3.38	0.4996
2.86	0.4979	3.04	0.4988	3.21	0.4993	3.39	0.4997
2.87	0.4979	3.05	0.4989	3.22	0.4994	3.40	0.4997
2.88	0.4980	3.06	0.4989	3.23	0.4994	3.48	0.4997
2.89	0.4981	3.07	0.4989	3.24	0.4994	3.49	0.4998
2.90	0.4981	3.08	0.4990	3.25	0.4994	3.50	0.4998
2.91	0.4982	3.09	0.4990	3.26	0.4994	3.61	0.4998
2.92	0.4982	3.10	0.4990	3.27	0.4995	3.62	0.4999
2.93	0.4983	3.11	0.4991	3.28	0.4995	3.63	0.4999
2.94	0.4984	3.12	0.4991	3.29	0.4995	3.89	0.4999
2.95	0.4984	3.13	0.4991	3.30	0.4995	2.90	0.5000
2.96	0.4985	3.14	0.4992	3.31	0.4995		
2.97	0.4985						

Ejercicios de repaso

1. Si la probabilidad de que una persona llegue a la edad de 83 años es de 0.002, calcula la probabilidad que de un total de 560 personas, exactamente tres personas lleguen a esa edad.

Sol. 0.0761

2. Si una distribución de Poisson es de $P(X)=\dfrac{(4.02)^x e^{-4.02}}{x!}$, calcula:

a) $P(0)$

b) $P(3)$

Sol. 0.01795, 0.19438

3. Calcula la probabilidad de obtener al menos dos caras en cuatro lanzamientos de una moneda.

Sol. 0.6875

4. Calcula el valor de la variable normalizada Z de una serie de números con media aritmética $\overline{X}=90$ desviación estándar $s=4$ para una observación de puntuación 126.

Sol. 9

5. Calcula el área bajo la curva normal entre $Z=1$ y $Z=1.5$, y expresa el resultado en la gráfica de una curva normal.

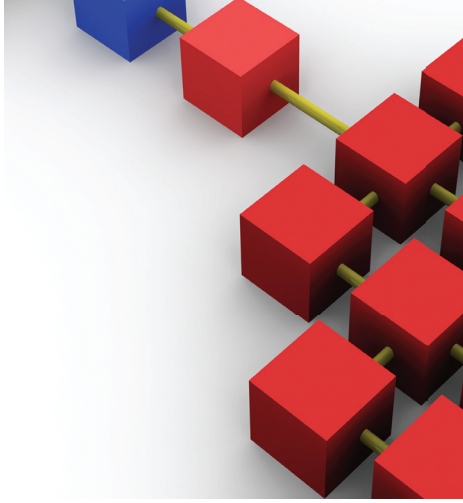
Sol. 0.0919

6. Calcula el área bajo la curva normal entre $Z=-1.5$ y $Z=1.2$, y expresa el resultado en la gráfica de una curva normal.

Sol. 0.8181

7. Calcula el valor de Z si el área que le corresponde en una curva normal entre 0 y Z es de 0.2517, y expresa el resultado en la gráfica de una curva normal.

Sol. 0.68



Capítulo 14

Correlación y regresión

Repaso de geometría analítica

Línea recta

En tu curso de *geometría analítica* aprendiste que la abscisa y la ordenada de un punto se llaman *coordenadas del punto* y se escriben como un par de números dentro de un paréntesis y separados por una coma. El primero de estos números representa siempre la abscisa y el segundo la ordenada. En general, para un punto A, cuya abscisa es x y la ordenada y, estas se designan con letras minúsculas con la notación $A(x, y)$. En estadística, sin embargo, nos referiremos a ellas como X y Y mayúsculas.

Conceptos clave

- Línea recta
- Correlación
- Coeficiente de correlación
- Diagrama de dispersión
- Regresión

Una ecuación de segundo grado con dos variables es del tipo:

$$Ax^2 + Cy^2 + Dx + Ey + F = 0$$

Si consideramos únicamente el polinomio:

$$Ax^2 + Cy^2 + Dx + Ey + F$$

Con $A = C = 0$ queda:

$$Dx + Ey + F$$

Como D, E y F son números reales cualquiera, podemos sustituirlos por las letras A, B y C para obtener:

$$Ax + By + C$$

Este polinomio es una función lineal y al igualarlo a cero es la ecuación de la recta en su forma general.

Ejemplos:

- **1.** $2x + y - 1 = 0$
- **2.** $6y - 4 = 0$
- **3.** $3x + 2 = 0$
- **4.** $y = 2x + 4$

Por este concepto se cita a la línea recta como una curva.

La expresión $y = bx + c$ tiene por gráfica la línea recta, donde los valores de b definen la dirección de la recta. Se le llama pendiente o coeficiente angular y los valores de c definen su posición.

Dos puntos (x_1, y_1) y (x_2, y_2) definen una recta y la pendiente m es:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

Ejemplos:

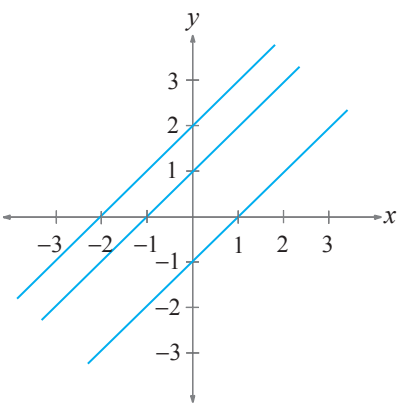
1. Las gráficas de las rectas:

L_1: $y = 3x + 2$

L_2: $y = 3x + 1$

L_3: $y = 3x - 1$

son paralelas entre sí por tener la misma pendiente $b = 3$ y forman una familia de rectas $y = kx + c$, que intersecan el eje de las y en (0, 2), (0, 1) y (0, −1), respectivamente.



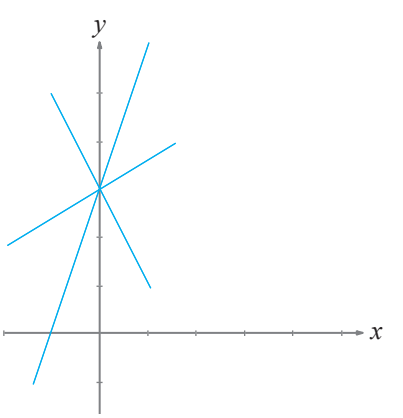
2. Las rectas:

L_1: $3x + 3$

L_2: $x + 3$

L_3: $-2x + 3$

no son paralelas, pero por tener el mismo coeficiente de posición $c = 3$ intersecan el eje de las y en el punto (0, 3) y sus pendientes son 3, 1, −2, respectivamente.



Las rectas $y = x$ y $y = -x$ son casos particulares de $y = bx + c$, donde el coeficiente de posición es $c = 0$ y pasan por el origen (0, 0). Las rectas que presentamos a continuación pasan por (0, 0) y forman un ángulo de 45º con los ejes del plano cartesiano.

$y = x$ con $b = 1$ y $c = 0$

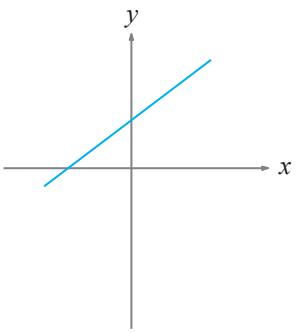
$y = -x$ con $b = -1$ y $c = 0$

A. La pendiente de toda recta paralela al eje x es 0.

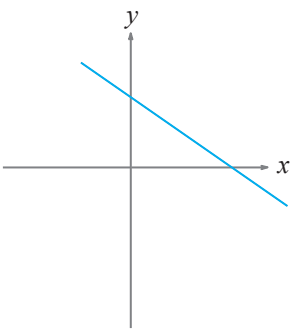
B. Una recta que forma un ángulo entre 0° y 90° tiene pendiente positiva.

C. Una recta paralela al eje y no tiene pendiente.

D. Si la recta forma un ángulo obtuso con el eje de las x, la pendiente es negativa.



Pendiente positiva



Pendiente negativa

Correlación

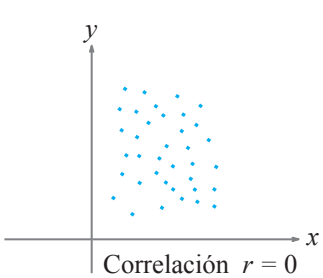
Hasta este punto hemos analizado diversos temas de estadística que incluyen una sola variable; sin embargo, es necesario investigar y estudiar la relación entre dos o más variables, como se muestra en los siguientes ejemplos:

- **1.** La que hay entre el tiempo que transcurre para que una persona se adapte a la oscuridad y la cantidad de azúcar en su sangre.
- **2.** Entre el peso de una persona, su edad y hábitos que lo hacen propenso a padecer una enfermedad.
- **3.** La que hay en el aprovechamiento entre los alumnos de enseñanza media superior y el tipo de escuela en que cursaron sus estudios básicos.
- **4.** La medida de las circunferencias de los círculos que depende de sus radios y que se han resuelto con la relación $C = 2\ \pi r$.

Coeficientes de correlación

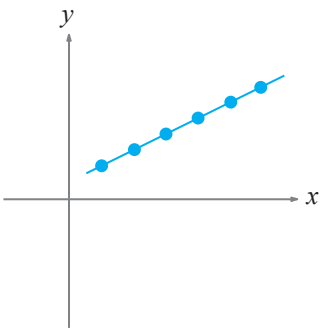
Son medidas que expresan la situación relativa de un número de sucesos respecto a dos variables. Son números cuyo valor varía entre los límites +1 y −1 y su magnitud se refiere al grado de asociación entre las variables.

Ejemplos:

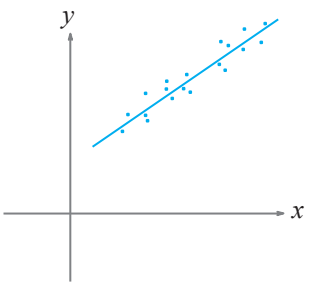


Correlación $r = 0$

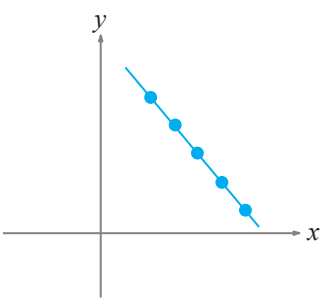
Variables no correlacionadas, no hay correlación, entonces $r = 0$



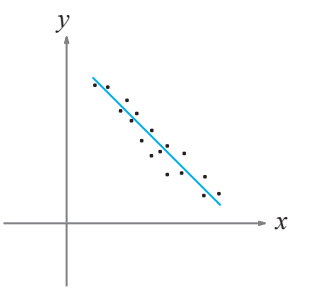
Correlación positiva
$r = 1$



Correlación positiva
$r = 0.8$



Correlación negativa
$r = -1$



Correlación negativa
$r = -0.8$

Conclusión:

- El valor $r = 0$ indica que no existe correlación entre las variables.
- Los valores +1 y −1 indican una correlación perfecta (lineal) positiva o negativa.

A continuación estudiaremos el cálculo de dos tipos de coeficientes de correlación lineal: la de Pearson y la de Spearman, que son las de uso más frecuente.

Para investigar la correlación entre dos variables se usan los **coeficientes de correlación**, que permiten expresar cuantitativamente el grado de relación que hay entre dos variables; por ejemplo, al estudiar la relación que se presenta entre los pesos de las personas que dependen, en cierta forma, de sus alturas, donde la muestra de n personas daría las alturas $X_1, X_2, \ldots, X_n$ y los pesos correspondientes $Y_1, Y_2, \ldots, Y_n$, datos que a continuación expresamos en un sistema de coordenadas rectangulares con los puntos $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.

Al conjunto de puntos ubicados se le llama **diagrama de dispersión**. Recuerda que las variables pueden ser independientes o dependientes. En la expresión $3x + 2y + 4$, las literales x y y son las variables.

Si los valores de una variable, por ejemplo y, dependen de los de otra variable, por ejemplo x, y realizadas las operaciones que se indiquen, si a cada valor de x le corresponde uno o más a y, decimos que hay una relación entre x y y, que x es la

variable independiente y y la dependiente; aunque si decidimos despejar x, entonces y sería la variable independiente y x la variable dependiente.

A menos de que indiquemos lo contrario, en este texto usaremos y como variable dependiente y x como variable independiente.

Coeficiente *r* de correlación lineal del producto momento (Pearson)

Este tipo de correlación se aplica para variables de intervalo o de razón y se calcula con la siguiente relación:

$$r = \frac{\sum xy}{\sqrt{\left(\sum x^2\right)\left(\sum y^2\right)}}, \text{ donde } x = X - \overline{X} \text{ y } y = Y - \overline{Y}$$

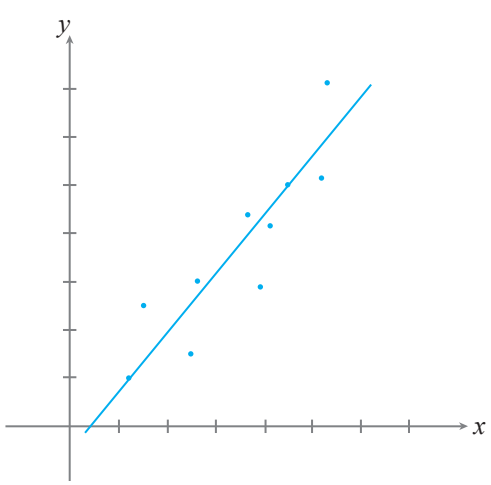
Observa que es necesario calcular la media aritmética de los valores de las abscisas de las coordenadas de los puntos, y el de las ordenadas de los mismos puntos.

Problema 1

Determina el coeficiente r de correlación lineal del producto-momento si las coordenadas de (X, Y) son: (1.5, 1), (2, 2.3), (2.5, 1.5), (3, 3), (4, 3), (4, 4.3), (4.5, 4.2), (5, 5.2), (6, 5.3), (6, 7.3).

Solución:

Diagrama de dispersión



Por el diagrama observamos que la correlación está cercana $a \pm 1$. Trazamos una recta con los puntos de coordenadas (1.5, 1) y (5, 5.2).

Calculamos los valores necesarios para sustituir en la relación:

$$r = \frac{\sum xy}{\sqrt{\left(\sum x^2\right)\left(\sum y^2\right)}}, \text{ donde } x = X - \overline{X} \text{ y } y = Y - \overline{Y}$$

Determinamos los valores de las medias aritméticas de $\overline{X}$ y de $\overline{Y}$:

$$\overline{X} = \frac{1.5+2+2.5+3+4+4+4.5+5+6+6}{10} = \frac{38.5}{10} = 3.85$$

$$\overline{Y} = \frac{1+2.3+1.5+3+3+4.3+4.2+5.2+5.3+7.3}{10} = \frac{37.1}{10} = 3.71$$

X	Y	$X-\overline{X}=x$	$Y-\overline{Y}=y$	x^2	xy	y^2
1.5	1.0	–2.35	–2.71	5.5	6.3	7.3
2.0	2.3	–1.85	–1.41	3.4	2.6	1.9
2.5	1.5	–1.35	–2.21	1.8	2.9	4.8
3.0	3.0	–0.85	–0.71	0.7	0.6	0.5
4.0	3.0	0.15	–0.71	0.0	0.1	0.5
4.0	4.3	0.15	0.59	0.1	0.0	0.3
4.5	4.2	0.65	0.49	0.4	0.3	0.2
5.0	5.2	1.15	1.49	1.3	1.7	2.2
6.0	5.3	2.15	1.59	4.6	3.4	2.5
6.0	7.3	2.15	3.59	4.6	7.7	12.8
38.5	37.1			22.3	25.6	33.0

En los resultados se tomó una cifra decimal.

Con:

$n = 10$

$\sum X = 38.5$

$\overline{X} = \frac{38.5}{10} = 3.85$

$\overline{Y} = \frac{37.1}{10} = 3.71$

$\sum x^2 = 22.3$

$\sum y^2 = 33.0$

$\sum xy = 25.6$

Sustituimos:

$$r = \frac{\sum xy}{\sqrt{\left(\sum x^2\right)\left(\sum y^2\right)}} = \frac{25.6}{\sqrt{(22.3)(33.0)}} = \frac{25.6}{\sqrt{735.9}} = \frac{25.6}{27.12}$$

$r = 0.94$

Coeficiente de correlación *r* por rangos de Spearman

Este coeficiente de correlación se aplica cuando uno o ambos conjuntos de medidas son ordinales, o sea, que los elementos de una serie o de ambas son posiciones. Por ejemplo, supongamos que una de las variables sea la estatura de 8 alumnos y la otra su edad.

Se calcula con la relación:

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

Si el conjunto de una de las variables no es ordinal, se asigna rango a sus puntuaciones. Además, es necesario calcular la diferencia D entre los rangos de sus puntuaciones.

Ejemplo:

Sea una de las coordenadas de una serie de datos la pareja (4, 7), el rango es $4 - 7 = -3$.

Problema 2

Calcula con el coeficiente de correlación r_s por rangos de Spearman el conjunto de datos siguientes:

(1, 3), (2, 1), (3, 2), (4, 5), (5, 4), (6, 8), (7, 4), (8, 8), (9, 7), que corresponden a nueve alumnos, donde el valor de la abscisa corresponde al lugar que ocupa cada uno en su grupo por su examen departamental de física y el de las ordenadas al examen departamental de literatura.

Solución:

En las coordenadas: (1, 3), (2, 1), (3, 2), (4, 5), (5, 4), (6, 8), (7, 4), (8, 8), (9, 7) calculamos el rango de cada una.

$1 - 3 = -2$	$4 - 5 = -1$	$7 - 4 = 3$
$2 - 1 = 1$	$5 - 4 = 1$	$8 - 8 = 0$
$3 - 2 = 1$	$6 - 8 = -2$	$9 - 7 = 2$

Calculamos los valores de D^2

D	–2	1	1	–1	1	–2	3	0	2
D^2	4	1	1	1	1	4	9	0	4

Con:

$n = 9$

$\sum D^2 = 25$

Sustituimos:

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} = 1 - \frac{6(25)}{9(81 - 1)} = 1 - \frac{150}{720} = 1 - 0.20$$

$r_s = 0.80$

Los coeficientes de correlación son sólo medidas de la covariación de las variables, pues la variación puede ser por causas que afectan a cada variable de la misma manera o de manera opuesta; o también, puede suceder que una de ellas sea la causa de la variación de la otra; todo esto es ajeno a la comprobación de la existencia de la correlación y del valor del coeficiente correspondiente.

Te recomendamos trazar siempre el diagrama de dispersión correspondiente, pues mostrará gráficamente si debe o no calcularse el valor de la correlación.

Regresión

En la **regresión** estudiamos la relación y la correlación entre dos variables, pero restringimos una de ellas, que permanece constante, para estudiar el comportamiento de la otra. Con esto podemos prever su valor en función de los valores que hayamos dado.

En todos los problemas de regresión hay una dependencia funcional entre las variables: si X es la variable independiente y Y la dependiente, se dice que hay una regresión de Y sobre X y se expresa en gráficas como líneas que pueden ser rectas o curvas. Por ahora nos referiremos únicamente a los ajustes que generan líneas rectas.

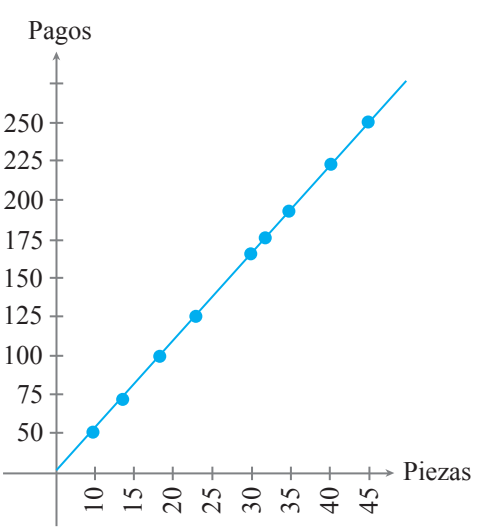
Problema 3

Una maquiladora paga a sus proveedores por pieza terminada y entregada, según el cuadro siguiente:

Piezas	Pago	Piezas	Pago
10	50	32	160
15	75	35	175
20	100	38	190
25	125	45	225

Traza el diagrama de dispersión, la gráfica y expresa la ecuación de la curva correspondiente.

Solución:



Ecuación de la recta: $Y = 5\,X$

Se le llama recta de regresión del pago sobre las piezas producidas.

Por ejemplo, 50 sobre 10 piezas, 75 sobre 15 piezas, condición que expresamos con números racionales: $\frac{50}{10}, \frac{75}{15}\ldots$

Problema 4

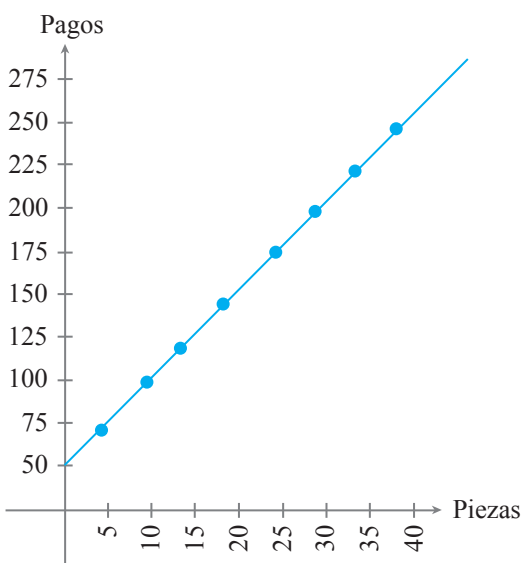
Los trabajadores que laboran en las proveedoras de la maquiladora que mencionamos en el problema anterior, piden a los dueños tener la opción de cotizar en el seguro social, lo que hace necesario modificar las condiciones de pago. Las partes convienen pagar un sueldo base equivalente a un salario mínimo, que asciende a 45 pesos y, sobre esta cantidad, continuar recibiendo 5 pesos por pieza entregada.

El cuadro de percepciones queda así:

Piezas	Pago	Piezas	Pago
10	95	32	205
15	120	35	220
20	145	38	235
25	170	45	270

Traza el diagrama de dispersión, la gráfica y expresa la ecuación de la curva correspondiente.

Solución:



La ecuación de la recta es $Y = bX + c$

Obtenemos el valor de la pendiente b:

$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{120 - 95}{15 - 10} = \frac{25}{5} = 5$ con $b = 5$ y obtenemos $y = 5X + c$. Para calcular el valor de c sustituimos en la ecuación, por ejemplo, con las coordenadas del punto (10, 95):

$95 = 5\ (10) + c$

$95 - 50 = c$

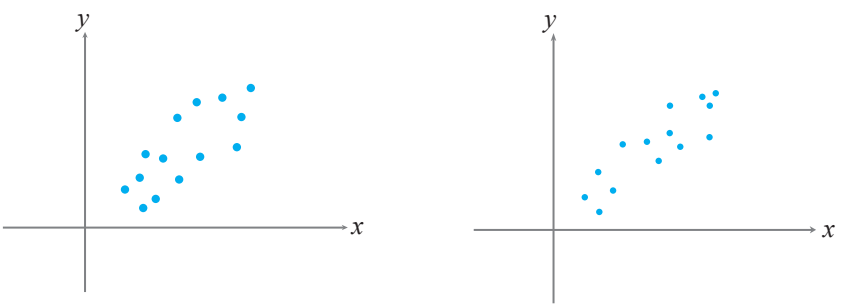
$c = 45$

Por lo tanto, $Y = 5X + 45$ es la solución.

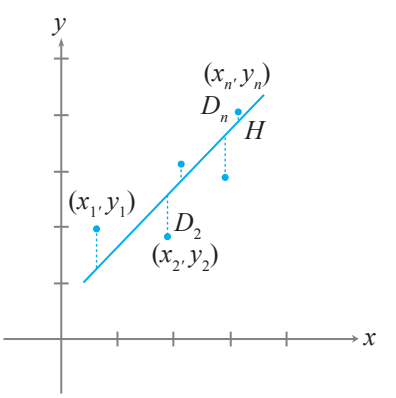
Las líneas de regresión permiten hacer lecturas fuera de los valores señalados; si se efectúa entre dos valores se dice que se ha hecho una **interpolación** entre esos valores; si leemos valores sobre la línea de la gráfica más allá de los valores experimentados habremos hecho una **extrapolación**.

Ajuste de curvas. Método de mínimos cuadrados

En otro tipo de problemas se obtienen casos como el siguiente: donde cada par de puntos define una recta y no existe alguna que pase por todos los puntos, pero aceptamos intuitivamente que sí hay una recta de mayor adecuación, de modo que sus desviaciones con respecto a los puntos sea lo menor posible.



Para evitar el juicio personal en la construcción de rectas de aproximación, se aplica el método de los mínimos cuadrados:



Consideremos los puntos representativos de unos datos y que éstos sean (X_1, Y_1), (X_2, Y_2)..., (X_n, Y_n). Si para un valor X, por ejemplo (X_1, Y_1) entre su valor y el correspondiente a la curva H, como se indica en la gráfica, habrá una diferencia D_1 se le llama desviación, misma que puede ser positiva, negativa o cero.

De igual forma para los valores $D_2, \ldots, D_n$.

Una medida de ajuste de la curva H a los puntos de los datos es el resultado de:

$D_1^2 + D_2^2 + \ldots D_n^2$ si éste es pequeño, el ajuste es bueno.

A todas las curvas de aproximación a un conjunto de datos puntuales que cumplen la condición citada se le llama la mejor curva de ajuste.

Recta de regresión de mínimos cuadrados

Si la ecuación de la recta que resulta por mínimos cuadrados de los puntos (X_1, Y_1), $(X_2, Y_2)\ldots, (X_n, Y_n)$ corresponde a la regresión de Y sobre X, su ecuación es $Y = bX + c$, que para facilitar su uso en estadística su expresa:

$$Y = c + bX \qquad (1)$$

La recta de regresión de X sobre Y se da con:

$$X = c + bY \qquad (2)$$

Las expresiones **(1)** y **(2)** también se pueden expresar, respectivamente, como:

$$y = \left(\frac{\Sigma xy}{\Sigma x^2} \right) x$$

$$x = \left(\frac{\Sigma xy}{y^2} \right) y$$

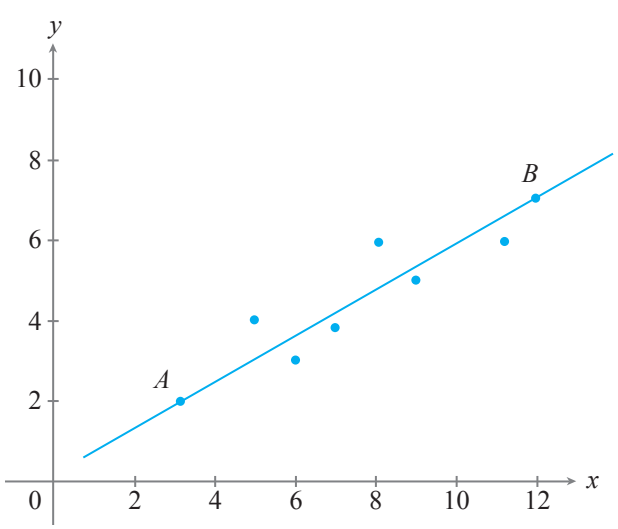
donde, $x = X - \overline{X}$ y $y = Y - \overline{Y}$

Las dos ecuaciones de regresión **(1)** y **(2)** serán idénticas, si y solamente si todos los puntos de dispersión están sobre la recta, entonces se dice que hay una correlación lineal perfecta entre X y Y, el valor $r = 1$ será para las dos.

Problema 5

En una investigación sobre costos, los pares de valores de (X, Y) son (3, 2), (5, 4), (6, 3), (7, 4), (8, 6), (9, 5), (11, 6) y (12, 6.8). Traza el diagrama de dispersión, la recta de regresión de Y sobre X que consideres por aproximación como la más adecuada y determina la ecuación de la recta por mínimos cuadrados que pase por lo puntos $A(3, 2)$ y $B(12, 6.8)$.

Solución:



Como la recta $Y = c + bX$ debe pasar por $A(3, 2)$ y $B(12, 6.8)$ para obtener la ecuación de la recta por mínimos cuadrados con las coordenadas de los puntos A y B planteamos un sistema lineal con dos incógnitas.

$2 = c + 3b$ **(1)**

$6.8 = c + 12b$ **(2)**

Lo resolvemos por adición y sustracción:

$$\begin{array}{r} -8 = -4c - \cancel{12b} \\ 6.8 = c + \cancel{12b} \\ \hline -1.2 = -3c \end{array}$$

$$c = \frac{1.2}{3} = 0.4$$

Sustituimos en **(1)**

$2 = c + 3b$

$2 = 0.4 + 3b$

$2 - 0.4 = 3b$

$3b = 1.6$

$$b = \frac{1.6}{3} = 0.53$$

Ecuación de la recta:

$Y = 0.4 + 0.53\,X$

Problema 6

Calcula el coeficiente r de correlación lineal si las coordenadas de (X, Y) son las mismas que se citaron en el problema anterior: (3, 2), (5, 4), (6, 3), (7, 4), (8, 6), (9, 5), (11, 6), (12, 6.8).

Considera el diagrama de dispersión y la gráfica de la recta obtenida en la resolución del problema citado.

Solución:

Calculamos los valores necesarios para sustituir en la relación

$r = \dfrac{\Sigma xy}{\sqrt{\left(\Sigma x^2\right)\left(\Sigma y^2\right)}}$, donde $x = X - \overline{X}$ y $y = Y - \overline{Y}$.

Calculamos ahora los valores de las medias aritméticas de $\overline{X}$ y $\overline{Y}$:

$$\overline{X} = \frac{3+5+6+7+8+9+11+12}{8} = \frac{61}{8} = 7.625$$

$$\overline{Y} = \frac{2+4+3+4+6+5+6+6.8}{8} = \frac{36.8}{8} = 4.6$$

X	Y	$X-\overline{X}=x$	$Y-\overline{Y}=y$	x^2	xy	y^2
3	2	−4.625	−2.6	21.3	12.025	6.7
5	4	−2.625	−0.6	6.8	1.575	0.3
6	3	−1.625	−1.6	2.6	2.6	2.5
7	4	−0.625	−0.6	0.3	0.375	0.3
8	6	0.375	1.4	0.1	0.525	1.9
9	5	1.375	0.4	1.8	0.55	0.1
11	6	3.375	1.4	11.3	4.725	1.9
12	6.8	4.375	2.2	19.1	9.625	4.8
61	36.8			63.3	32.000	18.5

Con:

$n=8$

$\sum X=61$

$\overline{X}=\frac{61}{8}=7.625$

$\sum \overline{Y}=\frac{36.8}{8}=4.6$

$\sum x^2=63.3$

$\sum y^2=18.5$

$\sum xy=32$

Sustituimos en:

$$r=\frac{\sum xy}{\sqrt{\left(\sum x^2\right)\left(\sum y^2\right)}}$$

$$r=\frac{32.000}{\sqrt{(63.3)(18.5)}}=\frac{32.000}{\sqrt{1171.05}}$$

$$r=\frac{32.000}{34.22}$$

$r=0.935$

Conclusión:

Al obtener el valor del coeficiente de correlación lineal *r* de Pearson, aceptamos que los resultados obtenidos están bien pues la distancia de la recta a las coordenadas de los puntos es menor a uno.

Ejercicios de repaso

1. Traza el diagrama de dispersión y calcula el rango de las coordenadas de los puntos (2.1, 2), (3.5, 3), (5, 3), (6, 4), (8, 5).

Sol. 0.1, 0.5, 2, 2, 3

2. Calcula la ecuación de la recta que pasa por los puntos de coordenadas (4, 3.5) y (10, 2.1).

Sol. $Y = 4.4 + 0.23\,X$

3. Dadas las coordenadas de los puntos (3, 2), (4, 4), (6, 4), (8, 5), (9, 7), (11, 7), (14, 8) calcula la media aritmética $\overline{X}$ de las abscisas y $\overline{Y}$ de las ordenadas.

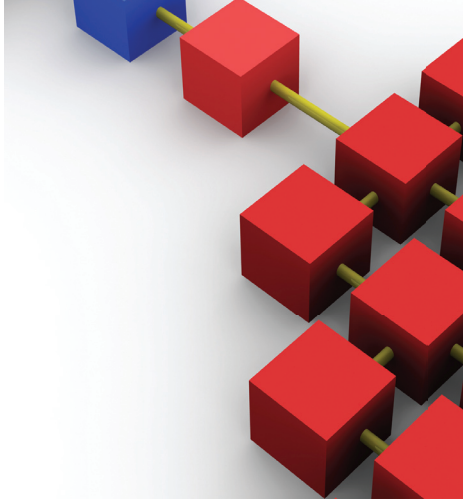
Sol. 7.85, 5.28

4. Calcula el valor del coeficiente *r* de correlación lineal de los puntos de coordenadas citados en el problema anterior; (3, 2), (4, 4), (6, 4), (8, 5), (9, 7), (11, 7), (14, 8).

Sol. 0.94

5. Determina la ecuación de la recta por mínimos cuadrados que pase por los puntos de coordenadas (3, 2) y (11,7) que corresponde a los datos del problema 3.

Sol. $Y = 0.125 + 0.625\,X$



Capítulo 15

Inferencia estadística. Conceptos básicos

Generalidades

Las diferencias más importantes entre estadística y probabilidad son:

- En la probabilidad se razona a partir de la población a la muestra.
- En la estadística, el razonamiento parte de la muestra para llegar al conocimiento de toda la población.

El estudio de una población, tomando como base las muestras, se llama **estadística inferencial** o **estadística inductiva**, algunos autores la citan como teoría de muestras.

La inferencia estadística trata de conocer, o explicar, el comportamiento de la población mediante los datos obtenidos de una muestra, e incluye:

- Muestreo
- Estimación puntual y por intervalos
- Prueba de hipótesis

Dado que no podemos estar absolutamente seguros de la veracidad de las inferencias obtenidas, las denominamos probabilidades.

Para predecir a partir de una muestra es necesario haberla seleccionado y recopilado cuidadosamente; si la muestra no se selecciona adecuadamente, es incorrecta o hay desviaciones en los datos, aún con cualquier tipo de análisis estadístico que se aplique, no se llegará a buenas conclusiones.

Conceptos clave

- Muestreo
- Estimación puntual y por intervalos
- Prueba de hipótesis
- Muestreo aleatorio con y sin reemplazo
- Muestreo por conglomerados
- Muestreo estratificado
- Muestreo sistemático
- Distribución de las medias de las muestras
- Estimador
- Estima del intervalo del parámetro
- Intervalos de confianza
- Hipótesis de trabajo
- Hipótesis de estadística
- Cursos de acción
- Ensayos de hipótesis

Muestreo

El proceso para obtener una muestra debe ser el más económico, el más rápido y el que asegure ser el más representativo de toda la población.

Al seleccionar una muestra debemos especificar claramente:

1. El método de selección de los individuos de la población y el tipo de muestra que se va a aplicar.
2. Tamaño de la muestra.
3. El grado de fiabilidad de las conclusiones que pensamos obtener.
4. Las características de la población de acuerdo a su grado de homogeneidad o heterogeneidad, respecto a la variable que se está analizando.

Puede ocurrir que una muestra represente a una población para determinadas variables y no sea representativa para otras.

Otros dos conceptos a tomar en consideración en el muestreo son:

Fracción de muestreo

Es el cociente que resulta de dividir el tamaño de la muestra y el tamaño de la población. Si se multiplica por 100, resulta ser el porcentaje de la población que representa la muestra.

Problema 1

En un pueblo de 17 500 habitantes se escoge una muestra de 850 personas para conocer el grado de seguridad ante delitos que las autoridades tratan de disminuir. Calcula la fracción de muestreo.

Solución:

$$\frac{850}{17\,500} = 0.0485$$

$0.04(100) = 4.8\ \%$

Se va a encuestar al 4% de la población.

Factor de elevación

Es el cociente entre el tamaño de la población y la muestra. El resultado representa el número de elementos que hay en la población por cada elemento de la muestra.

Problema 2

Con los datos del problema anterior calcula el factor de elevación.

Solución:

$$\frac{17\,500}{850} = 20.58$$

Cada persona de la muestra representa el 20.58 de la población (aproximadamente 21 personas).

Procedimientos de muestreo

Al decidir qué tipo de muestra o muestras se quieren, se tendrá en cuenta el número y características de la población.

Ejemplo:

Para expresar los colores que los adolescentes prefieren usar se podría usar la variable y = "Colores favoritos de los adolescentes"; sin embargo, los datos obtenidos no serían útiles para el estudio de las variables aleatorias = "Reacción de los propietarios de automóviles ante el pago de la tenencia".

En general, se consideran dos tipos de muestreo.

a) Probabilístico. Cada muestra tiene la misma probabilidad de ser elegida.

b) Intencional. La persona que obtiene la información es quien procura que la pregunta sea representativa de lo que se desea saber. Su representatividad es subjetiva.

El muestreo probabilístico puede ser:

- Aleatorio con y sin reemplazo
- Por conglomerados
- Estratificado
- Sistemático

Muestreo aleatorio con y sin reemplazo

Es aquel en que el proceso de selección de la muestra garantiza que todas las muestras posibles por obtener de la población pueden tener la misma probabilidad de ser elegidas.

Una vez que un elemento es seleccionado y las características del objeto de estudio son cuantificadas, vuelve a formar parte de la población y en consecuencia, puede volver a ser elegido. Éste es un muestreo aleatorio con reemplazo o reposición; se le cita con el nombre de aleatorio simple.

Si el elemento no vuelve a formar parte de la población, es un muestreo sin reposición o reemplazo. Se le identifica con el nombre de muestreo irrestrictamente aleatorio.

Los dos métodos son distintos, sin embargo, cuando el tamaño de la población es tan grande que puede considerarse como infinito; por ejemplo, la población del Distrito Federal mayor de 18 años, si se aplican los dos métodos no habrá diferencia en sus conclusiones.

Si la población es pequeña, por ejemplo, al revisar un pedido de mercancía, se recomienda el muestreo sin reemplazamientos para evitar que un elemento sea seleccionado más de una vez.

Muestreo por conglomerados

La población se divide en áreas que se llaman **conglomerados**, cada uno de éstos será lo más heterogéneo posible internamente y lo más homogéneo entre sí. A continuación se selecciona, al azar, uno o algunos conglomerados que forman la muestra.

Este método se utiliza cuando resulta muy costoso elaborar una lista completa de todos los elementos de la población; el inconveniente se presenta cuando los conglomerados no son homogéneos entre sí, ya que la muestra final puede no ser representativa de la población. Sin embargo, tiene la ventaja de simplificar el levantamiento de la población.

Ejemplo:

Sea la variable aleatoria *x*: "Intención de voto en las elecciones generales de una nación".

Para hacer el muestreo por conglomerados, el país se divide en regiones y éstas, a su vez, en ciudades con una población no mayor a 150 mil habitantes. Las localidades con más habitantes, se dividirán en municipios o barrios. La suma de estas divisiones representa a toda la población del país.

Las encuestas sobre tres candidatos pueden incluir, entre otras, preguntas como las siguientes:

1. De los candidatos propuestos, por cuál de ellos piensa votar. La respuesta podría ser condicionada por el temor al manifestar preferencia por un candidato.

2. De los problemas que sufren usted y sus familiares: inseguridad, falta de empleo, corrupción, encubrimiento, impunidad, ¿cuál de ellos le preocupa más y en qué orden le interesa sean resueltos?

3. De los gobiernos que hay en las diferentes entidades, ¿cuál o cuáles considera que están gobernados mejor? Explique brevemente por qué.

Se puede presentar el caso de que alguno de los partidos contendientes decida utilizar los resultados de encuestas anteriores y así favorecer su posición.

Muestreo estratificado

La población se divide en estratos homogéneos internamente y los más heterogéneos externamente entre sí. De cada estrato se selecciona un número de elementos proporcional al tamaño del estrato o según algún otro criterio (nivel económico, cultural, etcétera).

Si consideramos una población N y la dividimos en h subpoblaciones de tamaños N_1, ... N_h, éstas son disjuntas y cumplen $N_1 + N_2 + \ldots + N_k = N$.

Cada una de las subpoblaciones es un estrato.

Si necesitamos obtener una muestra de tamaño n de la población inicial, la obtenemos de cada estrato de manera que $n_1 + n_2 + \ldots n_h = n$.

Este método permite obtener las características de la información motivo de estudio y aumenta la precisión de las estimaciones sobre toda la población. En general, brinda mejores resultados que el muestreo aleatorio, mientras más diversos sean los estratos entre sí y sean más homogéneos internamente.

Las desventajas que presenta el muestro estratificado es que resulta difícil decidir a qué estrato asignar cada uno de los elementos de la población y cómo elegir el tamaño de la muestra de cada estrato para que el total sea n.

Problema 3

En una colonia con una población aproximada de 17 000 habitantes se sabe, según el censo reciente, que 7 800 son adultos, 2 950 de la tercera edad y 6 250 son niños. Calcula el tamaño de la muestra de cada estrato si se desea saber las preferencias de 300 personas en sus programas de televisión.

Solución:

Para las personas de la tercera edad:

$$300\left(\frac{2\,950}{17\,000}\right) = 300\ (0.17) = 51$$

Adultos:

$$300\left(\frac{7\,800}{17\,000}\right) = 300\ (0.45) = 135$$

Niños:

$$300\left(\frac{6\,250}{17\,000}\right) = 300\ (0.36) = 108$$

Nota: Se tomaron 2 cifras decimales en los valores de 0.17, 0.45 y 0.36. Se pueden redondear para que la suma sea de 300.

Muestreo sistemático

En este caso se divide la población en subconjuntos de tamaño; a continuación, se toma al azar un elemento del primer grupo que ocupa el lugar k y el resto de los elementos de la muestra ocupan los lugares.

$$k+\frac{N}{n}, k+2\frac{n}{N}, k, \ldots$$

Supongamos que la población es de N elementos ordenados y numerados del 1 hasta N, y queremos obtener una muestra del tamaño n. Dividimos la población en n subconjuntos, cada uno de ellos con $g = \frac{N}{n}$ elementos. Cada subconjunto constará de tantos elementos como indique el factor de elevación; además, despejando queda $N = ng$.

Se toma al azar un elemento de los enumerados desde 1, 2, … hasta $\frac{N}{n}$.

Si el resultado de $\frac{N}{n}$ no es entero, se redondea al entero menor. Esto puede producir una pequeña dificultad que no afecta y debe despreciarse cuando $n > 50$.

Este tipo de muestreo sistemático es semejante al aleatorio si los elementos se han numerado en forma aleatoria.

El muestreo es de aplicación fácil y se extiende la muestra a toda la población. La desventaja es que se presentan dificultades al tratar de calcular la varianza y su aumento si existe periodicidad en la numeración de los elementos.

Problema 4

Se aplicará una encuesta en una pequeña ciudad de 8 060 habitantes. Se seleccionará una muestra sistemática de 20 personas entre 1 200 padres de familia para conocer el grado de aceptación de la gestión administrativa de la ciudad por parte del presidente municipal.

Solución:

Calculamos el factor de elevación:

$$\frac{1\,200}{20} = 60$$

A continuación, seleccionamos un elemento al azar entre el 1 y el 60. Supongamos que elegimos el 27; así los demás elementos seleccionados son:

27, 87, 147, 207, 267, 327, 387, 447, 507, 567, 627, 687, 747, 807, 867, 927, 987, 1 047, 1 107, 1 167.

Se han seleccionado 20 personas, a las que les corresponden los números citados. Al número 27, que seleccionamos al azar, le sumamos 60 y continuamos del mismo modo hasta tener los 20:

(27, 27 + 60, 27 + 2(60), + 27 + 3(60),…)

Conclusión:

Para realizar encuestas más complejas, se utilizan los muestreos estratificados, conglomerados y aleatorios.

Problema 5

Se desea conocer la opinión sobre el desempeño de los funcionarios federales del gobierno de un país. Para lograr este objetivo, se solicita una proposición para la solución correspondiente.

Solución:

Una proposición sería la siguiente:

La población del país se dividirá en conglomerados: delegaciones en algunas ciudades o bien, municipios y barrios, los cuales pueden ser más o menos homogéneos internamente pero heterogéneos entre sí.

Posteriormente, estos conglomerados se clasifican en estratos homogéneos, por ejemplo en barrios. Cada uno de estos estratos, que son unidades primarias, se dividirá en nuevas unidades, por ejemplo, un número de manzanas o conjuntos habitacionales constituirán unidades secundarias.

Finalmente las muestras se tomarían así:

1. Se seleccionarían unas muestras estratificadas (al menos uno) de cada estrato.
2. De cada estrato seleccionado se eligen, al azar, varios bloques de casas.
3. Se seleccionará, al azar, una o varias casas dentro de los bloques citados.

Distribución de las medias de las muestras

Si tenemos una población y hemos obtenido una muestra (pueden ser varias), podemos obtener de ella otras muestras más pequeñas, señalando su tamaño. Con los valores obtenidos, podemos calcular la media aritmética, la varianza y la desviación estándar o típica. Este proceso se llama distribución de las medias de las muestras.

Problema 6

Una población incluye los números 2, 3, 6, 7, 8 y 10, considerando todas las muestras de tamaño 2 que pueden extraerse sin reemplazamiento.

Primero calcula la media aritmética, la varianza y la desviación estándar de la población citada. A continuación, tomando en cuenta un nuevo conjunto con todas las medias de las muestras de dos números, calcula el número de muestras y cuáles son sus elementos.

Solución:

Primera parte

- Media de la población:

$$\mu = \frac{2+3+6+7+8+10}{6} = \frac{36}{6} = 6$$

- Varianza:

$$\sigma^2 = \frac{(2-6)^2+(3-6)^2+(6-6)^2+(7-6)^2+(8-6)^2+(10-6)^2}{6}$$

$$= \frac{(-4)^2+(-3)^2+(0)^2+(1)^2+(2)^2+(4)^2}{6} = \frac{16+9+0+1+4+16}{6}$$

$$= \frac{46}{6} = 7.66 \text{ (se tomaron dos cifras decimales)}$$

- Desviación estándar o típica:

$$\sigma = \sqrt{7.66} = 2.76$$

Segunda parte

Número de muestras de tamaño 2 que se pueden formar.

$${}_6C_2 = \frac{6!}{2!(6-2)!} = \frac{6!}{2!4!} = \frac{6(5)(4!)}{2(1)(4!)} = \frac{30}{2} = 15$$

Como el número que se obtiene primero no se puede repetir, las muestras son:

(2, 3), (2, 6), (2, 7), (2, 8), (2, 10), (3, 6), (3, 7), (3, 8), (3, 10), (6, 7), (6, 8), (6, 10), (7, 8), (7, 10), (8, 10).

- Medias muestrales:

  2 + 3 = 5, 5 ÷ 2 − 2.5

  2 + 6 = 8, 8 ÷ 2 = 4

  2 + 7 = 9, 9 ÷ 2 = 4.5 y así con las demás, entonces las medias muestrales son:

  2.5, 4, 4.5, 5, 6, 4.5, 5, 5.5, 6.5, 6.5, 7, 8, 7.5, 8.5, 9

  Para citar la media, la varianza y la desviación estándar o típica se usan los símbolos siguientes:

  $\mu_{\bar{X}}, \sigma^2_{\bar{X}}, \sigma_{\bar{X}}$, respectivamente.

- Media de las medias muestrales:

$$\mu_{\bar{X}} = \frac{2.5+4+4.5+5+6+4.5+5+5.5+6.5+6.5+7+8+7.5+8.5+9}{15} =$$

$$= \frac{90}{15} = 6$$

- Varianza de las medias:

$$\sigma_{\bar{x}}^2 = \frac{(2.5-6.0)^2+(4-6.0)^2+(4.5-6.0)^2+(5-6.0)^2+(6-6.0)^2+(4.5-6.0)^2}{15}$$

$$+\frac{(5-6.0)^2+(5.5-6.0)^2+(6.5-6.0)^2+(6.5-6.0)^2+(7-6.0)^2+(8-6.0)^2}{15}$$

$$+\frac{(7.5-6.0)^2+(8.5-6.0)^2+(9-6.0)^2}{15}$$

$$=\frac{(-3.5)^2+(-2.0)^2+(-1.5)^2+(-1.0)^2+(-0.0)^2+(-1.5)^2+(-1.0)^2}{15}$$

$$+\frac{(-0.5)^2+(0.5)^2+(0.5)^2+(1.0)^2+(-2.0)^2+(1.5)^2+(2.5)^2+(3.0)^2}{15}$$

$$=\frac{12.25+4.00+2.25+1.00+0.00+2.25+1.00+0.25+0.25+0.25}{15}$$

$$+\frac{+1.00+4.00+2.25+6.25+9.00}{15}+\frac{46.00}{15}=3.06$$

- Desviación estándar o típica:

$$\sigma = \sqrt{3.06} = 1.74$$

La información obtenida con un buen muestreo, las medias, las varianzas y las desviaciones típicas, permiten establecer varias *hipótesis* que se valoran con los *estimadores*. Todo esto facilita tomar una buena decisión.

Estimación. Puntual y por intervalos

Una sola observación seleccionada al azar de una población no proporciona información suficiente para obtener conclusiones válidas.

Si la población es pequeña, como puede ser la producción de una fábrica en un día, una muestra que incluya varios elementos de la población será suficiente para hacer algunas observaciones. En casos en los que la población es grande, como la producción mensual de una fábrica, es necesario obtener varias muestras de la población.

Para ser justos en nuestras apreciaciones, recurrimos a la media aritmética $\bar{X} = \mu$ y a la varianza $s^2 = \sigma^2$ que se aplican a los elementos de cada muestra.

Ya definimos:

La media aritmética $\overline{X}$, como la suma de los valores de cierto número de cantidades dividido entre su número, es:

$$\overline{X} = \frac{\sum X}{N}$$

La varianza s^2 como la media aritmética de los cuadrados de desviaciones respecto a la media aritmética es:

$$s^2 = \frac{\sum X^2}{N}$$

Ejemplo:

Una cadena de tiendas de autoservicio quiere establecer una tienda en un barrio de la ciudad y necesita conocer el monto del ingreso medio de las familias; si decidieran preguntar a cada familia el resultado sería el más exacto pero el proceso sería costoso, lento y difícil, por ello, se decidieron a tomar una muestra de 475 familias donde calcularon la *media* aritmética $\overline{X}$ y el resultado lo utilizan como estimación de la media de los ingresos de toda la comunidad en estudio.

A los directivos también les interesa saber cómo se gastan estos ingresos (la dispersión) para así poder determinar los productos y precios de la mercancía que van a vender. La dispersión, es decir la desviación, puede ser estimada con la desviación típica $s = \sigma$ de la muestra.

De la muestra, o muestras obtenidas, podemos estimar cantidades desconocidas de la población, tales como la media aritmética $\overline{X} = \mu$, la varianza $s^2 = \sigma^2$ de ellas y, de ser necesario, se calcula también la desviación típica $s = \sigma$. A todos se les citan como parámetros poblacionales o simplemente parámetros y al calcular su valor como estadísticos muestrales o estadísticos.

Cuando se aplica un método estadístico para hacer la estimación de un parámetro, se le llama **estimador** y su valor que toma en una observación muestral es la **estimación puntual**.

El estimador es una variable aleatoria que depende únicamente de las características de la muestra. La estimación puntual, por el contrario, es un número, o sea el valor que el estimador tomó de la muestra. Se podrían utilizar como estimadores otros valores, por ejemplo, la mediana de la muestra, pero se obtienen datos más precisos si se usa la media aritmética.

Un buen estimador tiene las características siguientes:

1. Sin vicios o insesgado. La media de la distribución muestral del estadístico coincide con el parámetro poblacional desconocido.

   Un estimador es insesgado cuando el valor esperado de un método estadístico aplicado como estimador es igual al parámetro de la población que se va a estudiar.

   En estadística, no usamos expresiones como *preciso o exacto*. En su lugar decimos *estimador exacto* y *estimador insesgado*. En ocasiones, en lugar de decir *estimador preciso*, decimos *estimador de varianza mínima*.

2. Consistente. Cuando al aumentar el tamaño de la muestra, el valor de la media de la distribución muestral del estadístico muestral tiende al parámetro estimado.
3. Eficiente. Que sea el de menor varianza.
4. Suficiente. Si facilita toda la información sobre el parámetro que poseen los datos de la muestra.

Si nos referimos a un parámetro poblacional dado por un número, se dice que hay una estimación puntual, donde el valor de μ se estima con un valor de $\bar{X}$.

La media aritmética $\bar{X}$ es un buen estimador de μ; sin embargo, es necesario reconocer que hay un error entre ellas, el cual se expresa así:

$$e = \bar{X} - \mu$$

Si la estimación está dada por dos números, entre los cuales se encuentra el parámetro, se le llama **estima del intervalo del parámetro** y se expresa así:

$$a < \mu < b$$

La estimación por intervalos señala la exactitud o precisión de una estima y tiene un mayor uso que la puntual.

Los valores de a y b se obtienen de la observación de la muestra. Se usa con frecuencia y se les identifican como **intervalos de confianza**, donde a y b son los límites.

Ejemplos:

1. La altura media de un grupo de alumnos está entre 155 y 160 centímetros.

$$155 \text{ centímetros} < \mu < 160 \text{ centímetros}$$

2. La calificación de los alumnos de nuevo ingreso se encuentra entre 65 y 69 puntos.

$$65 < \mu < 69$$

3. El sueldo promedio de un trabajador calificado está entre 4 500 y 6 000 pesos.

$$4\,500 < \mu < 6\,000$$

4. Los transistores recibidos con defectos son del 4 al 6 por ciento.

$$4\% < \mu < 6\%$$

El concepto de intervalo de confianza permite que se asocie un valor de probabilidad al intervalo.

Comprobación de hipótesis (prueba de hipótesis)

La inferencia estadística es un proceso que se aplica a algo que se es capaz de conocer (cognoscitivo). Consiste en inferir una conclusión sobre una medida de una población (parámetro) con base en un estadístico obtenido de una muestra.

Si se establece una **hipótesis de trabajo** o **conjetura** fundamentada en la observación o en la experiencia personal, se denomina empírica. Si la hipótesis de trabajo se fundamenta en razonamientos propios de la disciplina dentro de la investigación que se hace, se le califica como lógica, que traducida al lenguaje de la estadística queda como **hipótesis estadística**.

En otras disciplinas, las hipótesis son **cursos de acción**.

Antes de tomar una decisión se analizan varias hipótesis o cursos de acción acerca de la población que se estudia.

En el supuesto de que se acepte una hipótesis acertada, si difiere de los resultados observados de aquellos que se esperaban o con la variación propia del muestreo, se dice que las diferencias son significativas y se está en condiciones de rechazarla de acuerdo con las evidencias obtenidas.

Los procedimientos que facilitan la decisión se citan como **ensayos de hipótesis**.

Errores de tipo I y tipo II

Si se rechaza una hipótesis cuando debería ser aceptada, se dice que se comete un error del tipo I. Si en cambio se acepta una hipótesis que debería ser rechazada, se comete un error del tipo II.

En ambos casos, se ha cometido un error al tomar la decisión equivocada. El ensayo de hipótesis se debe estructurar de manera que haga mínimos los errores citados, lo cual no es tan fácil, pues puede suceder que para un tamaño de muestra dado, el intento de disminuir un tipo de error vaya acompañado por el incremento de otro. Se valoran los riesgos porque una falla puede tener más importancia que otra; también se piensa que podemos poner una limitación al error de mayor importancia incrementando el tamaño de la muestra, lo cual puede ser posible o no.

Ejercicios de repaso

1. Una población incluye los números 1, 4, 6, 7, 12 y 15. Calcula la media aritmética, la varianza y la desviación estándar o típica de ella.

Sol. 7.50, 22.25, 4.71

2. Con la población de los números citados en el problema anterior, calcula cuántas muestras de tamaño 3 se pueden extraer sin reemplazamiento.

Sol. 20

3. En el almacén general de unas tiendas de autoservicio se recibe un lote de 8 750 llenadoras y se va a examinar aleatoriamente una muestra de 575 de ellas para verificar su funcionamiento. Calcula la fracción del muestreo.

Sol. 6%

4. Con los datos del problema anterior, calcula el factor de elevación.

Sol. 15.21

5. Tenemos una población con los números 4, 7, 9, 12, 18. Determina el número de muestras de tamaño 2 que se pueden obtener sin reemplazamiento, cuáles son y calcula la media aritmética de cada una.

Sol. 10

(4, 7), (4, 9), (4, 12), (4, 18), (7, 9), (7, 12), (7, 18), (9, 12), (9, 18), (12, 18), 5.5, 6.5, 8, 11, 8, 9.5, 12.5, 10.5. 13.5. 15.0

SERIE UNIVERSITARIA

PATRIA

Teoría, ejemplos y problemas

CD interactivo en esta edición

PROBABILIDAD y ESTADÍSTICA

Victor M. Alvarado Verdin

GRUPO EDITORIAL PATRIA

439

PROBABILIDAD Y ESTADÍSTICA

Estadísticos e-Books & Papers

Estadísticos e-Books & Papers

PROBABILIDAD Y ESTADÍSTICA

Víctor Manuel Alvarado Verdín

PRIMERA EDICIÓN EBOOK
MÉXICO, 2014

GRUPO EDITORIAL PATRIA

442
Estadísticos e-Books & Papers

Para establecer comunicación con nosotros puede hacerlo por:

correo:
Renacimiento 180, Col. San Juan Tlihuaca, Azcapotzalco, 02400, México, D.F.

fax pedidos:
(01 55) 5354 9109 • 5354 9102

e-mail:
info@editorialpatria.com.mx

home page:
www.editorialpatria.com.mx

Dirección editorial: Javier Enrique Callejas
Coordinadora editorial: Estela Delfín Ramírez
Supervisor de prepensa: Gerardo Briones González
Diseño de portada: Juan Bernardo Rosado Solís
Fotografías: © Thinkstockphoto

Revisión técnica:
Carlos Gerardo Velázquez Rodríguez
UVM-Coacalco/ESIQIE-Instituto Politécnico Nacional
Ana Elizabeth García Hernández
Instituto Politécnico Nacional

Probabilidad y estadística

Derechos reservados:
© 2014, Víctor Manuel Alvarado Verdín
© 2014, Grupo Editorial Patria, S.A. de C.V.
Renacimiento 180, Colonia San Juan Tlihuaca,
Delegación Azcapotzalco, Código Postal 02400, México, D.F.

Miembro de la Cámara Nacional de la Industria Editorial Mexicana
Registro núm. 43

ISBN ebook: 978-607-438-930-2

Queda prohibida la reproducción o transmisión total o parcial del contenido de la presente obra en cualesquiera formas, sean electrónicas o mecánicas, sin el consentimiento previo y por escrito del editor.

Impreso en México
Printed in Mexico

Primera edición ebook: 2014

Estadísticos e-Books & Papers

Agradecimientos

A familiares y amigos por el apoyo prestado durante el desarrollo de este contenido; espero que disculpen los momentos de abstracción.

A Tere, Geo, Osiris y Doris por su apoyo al recopilar y ordenar las notas y apuntes de quien suscribe.

A mi editora Estela Delfín R., por su orientación y paciencia durante el desarrollo del texto.

A la Dra(c). Sánchez Arrioja por su paciencia y apoyo durante el desarrollo del libro.

Estadísticos e-Books & Papers

Estadísticos e-Books & Papers

Prefacio

A últimas fechas, tanto la Probabilidad como la Estadística se han erigido como las materias o asignaturas académicas que ofrecen las herramientas de orden objetivo, a fin de generar procesos de pensamiento lógico-racionales basados en el análisis de la información cuantitativa relativa a todo tipo de actividades y fenómenos desarrollados por personas y organizaciones.

El panorama propuesto exige a estudiantes y académicos mantenerse familiarizados con los principios, métodos y técnicas matemáticos que les faciliten organizar, sintetizar, computar y analizar información cuantitativa, con el objetivo de poder tomar las decisiones más adecuadas, así como para la optimización del desarrollo de aplicaciones con propósitos específicos.

En atención a lo expuesto antes, la estructura de los contenidos de este libro ha sido diseñada con el propósito de:

- Exponer de manera resumida los principales temas relativos a la Probabilidad y a la Estadística, en términos y palabras sencillos y simples, evitando un lenguaje matemático complejo.
- Destacar notas al margen donde se detallan aquellos conceptos que se consideran trascendentes, a fin de que el estudiante se familiarice con estos, relacionándolos con los desarrollos temáticos.
- Desarrollar los procesos de solución de problemas basados en aplicaciones de orden práctico, los cuales permitan al estudiante repasar los métodos de cálculo, con el objetivo de que este refuerce su entendimiento temático.
- Proponer un amplio catálogo de problemas resueltos y para resolver; así, al final de cada unidad se proporcionan ejercicios y problemas cuyos planteamientos exponen retos teórico-conceptuales para quienes intentan resolverlos.

En términos generales, este libro se considera un texto útil para el complemento de las exposiciones teóricas dictadas en el aula, quedando en manos de los estudiantes y los académicos la disertación temática y la orientación práctica del presente contenido.

El Autor

446
Estadísticos e-Books & Papers

Estadísticos e-Books & Papers

Contenido

Unidad 1 La estadística y la estadística descriptiva 1

1.1 Introducción 2
1.2 Definición de estadística 2
1.3 Clasificación de la estadística 2
1.4 La estadística descriptiva 2
1.5 Análisis de datos por estadística descriptiva para datos no agrupados 3
1.6 Análisis de estadística descriptiva para datos agrupados 8
1.7 Gráficas descriptivas: histograma, polígono de frecuencias y ojiva 13
1.8 Medidas de tendencia central para datos agrupados 20
1.9 Determinación de la ubicación y el valor de la moda de manera gráfica mediante el histograma y el polígono de frecuencias 22
1.10 Determinación de la ubicación y el valor de la mediana de manera gráfica mediante la ojiva 24
1.11 Cálculo de cuantiles para datos agrupados 25
1.12 Momentos estadísticos 26
1.13 Determinación del número de intervalos de clase 29
1.14 La media geométrica 31

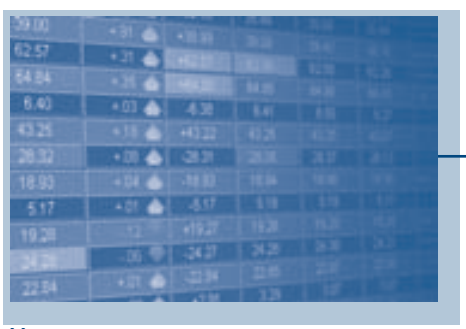
Problemas para resolver 33
Problemas reto 35
Referencias 37
Direcciones electrónicas 37

Estadísticos e-Books & Papers

Unidad 2 Teoría de la probabilidad y distribuciones de probabilidad 39

2.1 Introducción 40
2.2 Concepto de probabilidad 40
2.3 Los espacios muestrales y la teoría de conjuntos 40
2.4 Probabilidad clásica 42
2.5 Probabilidad de eventos mutuamente exclusivos o excluyentes 43
2.6 Probabilidad de eventos comunes 44
2.7 Probabilidad de eventos simultáneos o sucesivos 45
2.8 Los diagramas de árbol 46
2.9 Eventos consecutivos con y sin reemplazo 47
2.10 Probabilidad condicional 48
2.11 Teorema de Bayes 48
2.12 El principio de multiplicación 50
2.13 Permutaciones 51
2.14 Combinaciones 52
2.15 Variables aleatorias 53
2.16 Distribuciones de probabilidad 53
2.17 Curva de distribución de probabilidad discreta 58
2.18 Valor esperado de una variable aleatoria discreta o esperanza matemática 59
2.19 Distribuciones de probabilidad continuas 60
2.20 Teorema del límite central 63

Problemas para resolver 64
Problemas reto 68
Referencias 69
Direcciones electrónicas 69



Unidad 3 Estadística inferencial 71

3.1 Introducción 72
3.2 Concepto y propósito de la estadística inferencial 72

Estadísticos e-Books & Papers

3.3 Estimadores 73
3.4 Distribución de las medias muestrales 75
3.5 Distribución muestral de las proporciones 80
3.6 Intervalos de confianza 83
3.7 Determinación del tamaño muestral para estimar una media poblacional 86
3.8 Determinación del tamaño muestral para estimar la proporción poblacional 88
3.9 Grados de libertad 89
3.10 Intervalos de confianza para muestras pequeñas
Problemas para resolver 92
Problemas reto 95
Referencias 95
Direcciones electrónicas 96

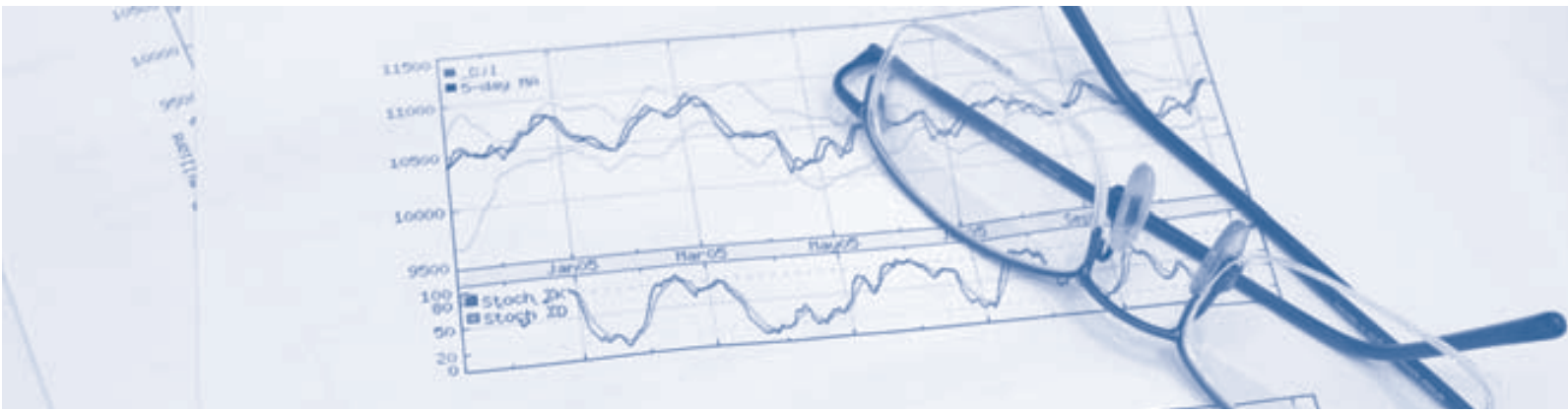
Unidad 4 **Análisis estadístico de experimentos** 97

4.1 Introducción 98
4.2 Concepto de experimento 98
4.3 Hipótesis estadísticas 98
4.4 Aplicación de la distribución *t* de Student en el análisis de experimentos 99
4.5 Estudios comparativos simples 99
4.6 Estudios comparativos basados en pruebas de hipótesis sustentados en el análisis de la varianza (ANOVA) de un factor con una muestra por grupo 102
4.7 Estudios comparativos basados en ANOVA de dos factores 105
4.8 Estudios comparativos basados en ANOVA de dos factores de varias muestras por grupo 109
4.9 Discriminantes para pruebas de hipótesis basadas en ANOVA 114
4.10 Método de discriminación de la *R* de Duncan 120
Problemas para resolver 123
Problema reto 127
Referencias 128
Direcciones electrónicas 128

Estadísticos e-Books & Papers

Estadísticos e-Books & Papers

UNIDAD 1



La estadística y la estadística descriptiva

OBJETIVOS

- Entender la importancia de la estadística como la herramienta que facilita el análisis y la interpretación de la información cuantitativa relacionada con los fenómenos de interés que acontecen en la realidad.
- Aplicar el proceso de análisis de la estadística descriptiva.
- Entender la diferencia entre datos agrupados y no agrupados.
- Entender el concepto de medidas de tendencia central.
- Entender el concepto de medidas de dispersión.
- Entender el concepto de medidas de posición.
- Desarrollar los elementos gráficos básicos de la estadística descriptiva.
- Desarrollar el proceso de análisis por momentos estadísticos.

¿QUÉ SABES?

- ¿Qué es la estadística y para qué sirve?
- ¿Cuál es la diferencia entre la estadística descriptiva y la inferencial?
- ¿Cuál es la diferencia entre los datos agrupados y los no agrupados?
- ¿Cuáles son las principales medidas de tendencia central y qué significan?
- ¿Qué son los intervalos de clase?
- ¿Cuántos tipos de frecuencia existen?
- ¿Cuáles son las principales gráficas del análisis de la estadística descriptiva?
- ¿Qué es una curva de distribución de frecuencias?
- ¿Cuáles son los principales cuantiles?
- ¿Para qué sirven los momentos estadísticos?

Estadísticos e-Books & Papers

1.1 Introducción

La estadística proporciona la metodología para el tratamiento de la información cuantitativa relacionada con hechos y sucesos de la realidad, fundamentada en la organización y cómputo de la misma, con el propósito de generar conocimiento sobre los mismos.

1.2 Definición de estadística

Por estadística debe entenderse el proceso sistemático aplicado al análisis y la interpretación de numéricos con la intención de comprender los hechos de la realidad pudiendo apoyar la toma de decisiones racionales.

La definición anterior expone que la estadística es un conjunto de actividades interrelacionadas, las cuales se desarrollan bajo principios bien definidos y en estricto orden, a efecto de coleccionar, organizar, computar, presentar e interpretar los conjuntos de cifras numéricas (datos) que guardan relaciones significativas con un fenómeno en particular.

Asimismo, a los conjuntos de datos se les denomina colección de datos, mismos que se pueden clasificar en:

- **No agrupados**, cuando los datos que las conforman a lo más guardan un orden secuencial de acuerdo con su valor.
- **Agrupados**, cuando los datos que las conforman han sido catalogados dentro de un grupo de rangos denominados intervalos de clase en atención a que representan al grupo de rangos en los que se puede subdividir la colección, permitiendo clasificar los datos de acuerdo con su valor dentro de los mismos.

1.3 Clasificación de la estadística

La estadística, en atención a su aplicación y propósitos, se clasifica en:

Descriptiva, la cual tiene por objeto organizar y presentar conjuntos de datos numéricos con la intención de facilitar el análisis y la caracterización de un fenómeno.

Inferencial, la cual tiene como finalidad la validación de los parámetros de una población mediante los estadísticos de una o varias muestras.

1.4 La estadística descriptiva

El análisis por estadística descriptiva se fundamenta en el cálculo de las llamadas medidas descriptivas, las cuales recapitulan la información de una colección de datos permitiendo describir el comportamiento de un fenómeno.

Las medidas descriptivas de uso común se clasifican de la siguiente forma:

a) **Medidas de centralización.** Son el grupo de valores más representativos de una colección ordenada de datos, los cuales tienden a ubicarse al centro de la misma. Entre las medidas más representativas se pueden citar:

   - La media.
   - La mediana.
   - La moda.
   - El promedio ponderado.
   - El promedio geométrico.

b) **Medidas de dispersión.** Señalan qué tan alejados están los valores de una colección de datos con respecto a un valor de centralización, que por lo general es la media. Entre las medidas más comunes se encuentran:

Estadísticos e-Books & Papers

- El rango.
- La varianza.
- La desviación estándar.
- El coeficiente de variación.

c) **Medidas de posición o cuantiles.** Son los valores que permiten dividir la colección ordenada de datos en partes iguales con el mismo número de datos en cada segmento. Los cuantiles más comunes son:

- Los percentiles, los cuales dividen la colección en 100 partes iguales, considerando que existen 99 percentiles (P_1, P_2, P_3… P_{99}).
- Los deciles, los cuales dividen la colección en 10 partes iguales, considerando que existen 9 deciles (D_1, D_2, D_3… D_9).
- Los cuartiles, los cuales dividen la colección en 4 partes iguales, considerando que existen 3 cuartiles (Q_1, Q_2 y Q_3).

Es de notar que el P_{50}, el D_5 y el Q_2 representan el valor de la mediana.

d) **Medidas de forma.** Son los valores que permiten establecer cómo están distribuidos los valores de una colección de datos. Las medidas principales son:

- Sesgo o asimetría.
- Apuntamiento.

Alerta

Las medidas descriptivas son de centralización y de forma.

1.5 Análisis de datos por estadística descriptiva para datos no agrupados

Este tipo de análisis se recomienda cuando el número de datos que estructuran una colección de datos permite su manejo y cómputo de manera ágil.

Para que las cifras ofrezcan un significado es conveniente ordenarlas, sugiriendo en este caso de menor a mayor, de acuerdo con sus valores.

Alerta

Los datos no agrupados no se encuentran clasificados por categorías o intervalos.

Problema resuelto

Procede a ordenar de menor a mayor la siguiente colección de datos.

18 5 11 52 35 52 72

Solución

Ordenando de menor a mayor:

5 11 18 35 52 52 72

Ordenar los datos permite contar con una mejor perspectiva de los mismos, pudiendo establecer las diferencias entre los diferentes valores.

■ Medidas de tendencia central para datos no agrupados

Debe recordarse que de manera básica son tres las medidas de tendencia central: media, mediana y moda.

a) **Determinación del valor de la media**

De manera concreta, la media o promedio representa el valor más representativo de una colección de datos tendiendo a ubicarse al centro de la misma, cuyo valor permite establecer un equilibrio

Estadísticos e-Books & Papers

en cuanto a las diferencias existentes con el resto de los valores.

Matemáticamente, la media queda definida como

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

donde

$\overline{x}_i$ = Representa el valor i ($i = 1, 2, 3, \ldots n$).

n = Número total de datos de la colección.

Problema resuelto

Considerando la colección

5 11 18 35 52 52 72

Determina el valor de la media.

Solución

El valor de la media del ejemplo en cuestión es:

$$\overline{X} = \frac{5+11+18+35+52+52+72}{7} = 35$$

Nótese que la media dentro de la colección ordenada se ubica exactamente en el centro:

5 11 18 35 52 52 72

↑

$\overline{X}$

Alerta

La media es el valor más representativo de una colección de datos.

b) Determinación del valor de la mediana

La siguiente medida de tendencia central es la mediana (M_d), la cual representa el valor central de la colección.

La determinación de la mediana debe cumplir con las siguientes reglas:

I. Si el número de datos de la colección es impar, el valor de la mediana es el valor central de la misma dividiendo en dos segmentos iguales a la colección.

II. Si el número de datos de la colección es par, el valor de la mediana es el promedio aritmético de los valores centrales.

Alerta

La mediana es el valor central de una colección de datos.

Problema resuelto

Considerando la colección

5 11 18 35 52 52 72

Determina el valor de la mediana.

Solución

En el caso de la colección en análisis, esta cuenta con un número impar de datos, por lo que el valor de la mediana es el que cumple con la regla I, antes mencionada, coincidiendo con la posición y el valor de la media.

Estadísticos e-Books & Papers

Solución *(continuación)*

5 11 18 35 52 52 72

$\bar{X}$

M_d

c) Determinación del valor de la moda

En el caso de la moda (M_o), es el valor o valores que tienen la mayor frecuencia, o sea, son los que más se repiten.

Con referencia a lo anterior, debe considerarse que en una colección de datos puede haber más de una moda, por lo que una colección puede ser:

- **Modal**, cuando cuenta con un valor con mayor frecuencia.
- **Bimodal**, cuando cuenta con dos valores con la misma frecuencia.
- **Multimodal**, cuando cuenta con más de dos valores con la misma frecuencia.

Problema resuelto

Considerando la siguiente colección

5 11 18 35 52 52 72

Determina el valor de la moda.

Solución

Puede observarse que la colección es modal ya que el valor que más se repite es el 52, considerando que cuenta con una frecuencia con valor 2.

5 11 18 35 52 52 72

M_o

Alerta

La moda es el valor o valores que más se repiten en una colección de datos; una colección puede tener más de una moda.

Medidas de dispersión para datos no agrupados

La determinación de los valores de las medidas de dispersión es relativamente simple ya que el valor de referencia es la media de la colección, por lo que se deben establecer las diferencias o desviaciones existentes entre los distintos valores de la colección con respecto a ella.

Si esto se realiza, se encuentra que el valor promedio es cero, ya que la media equilibra las desviaciones tanto por arriba como por debajo de la misma, tal como se muestra en el siguiente problema resuelto.

Problema resuelto

Considerando la siguiente colección

5 11 18 35 52 52 72

Comprueba que la suma de las diferencias con respecto a la media es cero.

Estadísticos e-Books & Papers

Alerta

Las medidas de dispersión permiten establecer el grado de alejamiento entre un valor específico en relación con el resto de los demás datos.

Solución

Tabla 1.1

x	$x - \bar{X}$
5	$5 - 35 = -30$
11	$11 - 35 = -24$
18	$18 - 35 = -17$
35	$35 - 35 = 0$
52	$52 - 35 = 17$
52	$52 - 35 = 17$
72	$72 - 35 = 37$
$\Sigma =$	0

Por tanto, no se puede establecer la diferencia o desviación promedio de los valores de la colección con respecto a la media, pero para evitar esto se procede a elevar las diferencias al cuadrado a efecto de evitar los números negativos y obtener un promedio de las desviaciones; a este parámetro se le denomina varianza (S^2), pero ha de observarse que la varianza expone el promedio de las desviaciones al cuadrado por lo que un valor más significativo lo propone la desviación estándar, la cual es la raíz cuadrada del valor de la varianza. Para ejemplificar lo anterior obsérvese el siguiente problema resuelto.

Problema resuelto

Considerando la colección

5 11 18 35 52 52 72

Determina el valor de la varianza y de la desviación estándar.

Solución

Tabla 1.2

x	$(x - \bar{X})^2$
5	$(5 - 35)^2 = 900$
11	$(11 - 35)^2 = 576$
18	$(18 - 35)^2 = 289$
35	$(35 - 35)^2 = 0$
52	$(52 - 35)^2 = 289$
52	$(52 - 35)^2 = 289$
72	$(72 - 35)^2 = 1369$
$\Sigma =$	3 712

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} = 530.29$$

Sin embargo, la varianza expone un promedio de cuadrados, por lo que se procede a calcular la raíz cuadrada para obtener un valor más significativo, al cual se le denomina desviación estándar.

$$S = \sqrt{S^2} = 23.03$$

Estadísticos e-Books & Papers

Medidas de posición para datos no agrupados

En el caso de las colecciones de datos no agrupados, por lo regular se procede a determinar los valores de las principales medidas de posición como lo son los cuartiles. Para determinar estos valores se utiliza la siguiente fórmula:

$$PS_p = (n+1)\left(\frac{P}{100}\right)$$

Donde

PS_p = Posición del percentil P.

n = Número de elementos de la colección.

La fórmula anterior arroja la posición del percentil de interés, por lo que se deberá determinar el valor del mismo mediante diferencias y proporciones, como se explicará más adelante. Para ejemplificar lo anterior se formula el siguiente problema.

Alerta

Las medidas de posición establecen los valores que permiten dividir la colección en segmentos iguales.

Problema resuelto

Considerando la colección en análisis, determina los valores del primer y tercer cuartiles, así como del 80avo percentil.

5	11	18	35	52	52	72

Solución

5	11	18	35	52	52	72
1	2	3	4	5	6	7

En la primera línea de la serie anterior se muestra de manera ordenada la colección y en la segunda línea la posición de cada valor dentro del arreglo ordenado.

Para determinar el primer cuartil se procede como sigue:

$P = 25$

$n = 7$

por tanto,

$$PS_{25} = (7+1)\left(\frac{25}{100}\right) = 2$$

Donde se interpreta que el valor del primer cuartil corresponde al del dato ubicado en la posición 2 de la colección ordenada, que en este caso es $Q_1 = PS_{25} = 11$.

Para determinar el tercer cuartil se procede de igual forma si $P = 75$,

$$PS_{75} = (7+1)\left(\frac{75}{100}\right) = 6$$

Donde se interpreta que el valor del tercer cuartil corresponde al del dato ubicado en la posición 6 de la colección ordenada, que en este caso es $Q_3 = PS_{75} = 52$.

En el caso del 80avo percentil $P = 80$:

$$PS_{80} = (7+1)\left(\frac{80}{100}\right) = 6.4$$

El valor de $P = 80$ se encuentra en la posición 6.4, la cual no se encuentra de manera explícita, por lo que se procede a determinar la distancia entre los valores de las posiciones 6 y 7,

$$\text{Dif} = 72 - 52 = 20$$

Estadísticos e-Books & Papers

Solución *(continuación)*

Entonces se puede señalar que el valor en cuestión se encuentra a 40% de la distancia a partir del valor de la posición 6:

$$X = \text{Dif} \times 40\% = 20 \times 0.4 = 8$$

Por tanto,

$$P_{80} = 52 + X = 60$$

Para completar los referentes sobre las colecciones no agrupadas considérese el caso de una colección de datos con número de elementos par tal como se expone en el siguiente problema resuelto.

Problema resuelto

Considerando la siguiente colección de datos, procede a determinar los valores y la ubicación de las medidas de tendencia central, así como de las medidas de dispersión.

23 40 54 69 69 69

Solución

Se procede a calcular las medidas de tendencia central, comenzando por la media ($\overline{X}$):

$$\overline{X} = \frac{\Sigma_{i=1}^{n} x_i}{n} = \frac{23 + 40 + 54 + 69 + 69 + 69}{6} = 54$$

En este caso en particular el valor de la mediana (M_d) se determina a través del promedio de los datos centrales, mismos que permiten dividir la colección en dos segmentos iguales.

$$23 \quad 40 \quad \underline{54 \quad 69} \quad 69 \quad 69$$

$$M_d = \frac{54 + 69}{2} = 61.5$$

Se observa que el valor de la moda (M_o) es 69 debido a que el valor de su frecuencia es de 3.

$$M_o = 69$$

Adicionalmente, se calcula el valor de las medidas de dispersión, comenzando con el rango.

$$R = \textit{Valor máximo} - \textit{Valor mínimo} = 69 - 23 = 46$$

Donde el valor de la varianza (S^2) es:

$$S^2 = \frac{\Sigma_{i=1}^{n} (x_i - \overline{X})^2}{n} = 305.33$$

En consecuencia el valor de la desviación estándar es:

$$S = \sqrt{S^2} = 17.47$$

1.6 Análisis de estadística descriptiva para datos agrupados

Cuando el número de datos de una colección es grande puede generar dificultades para su manejo y cómputo, por lo que se aconseja organizarlos a través de intervalos de clase. Los intervalos de clase, como se mencionó antes, son un conjunto de clases en las que se puede dividir el rango de una colección de datos.

Estadísticos e-Books & Papers

La tabla de frecuencias

Tabla 1.3 Tabla de frecuencias

Intervalos nominales			*f*
LI		LS	
5	-	9	3
10	-	14	7
15	-	19	14
20	-	24	4
25	-	29	2
$\sum f$	=	n =	30

La tabla anterior se denomina tabla de frecuencias y en ella se observa que la primera columna es la de los intervalos nominales, los cuales son los rangos que permiten clasificar los datos de la colección de manera *a priori*, donde cada intervalo de clase cuenta con un límite inferior (LI) y un límite superior (LS); la segunda columna es la de las frecuencias, cuyos valores representan el número de datos de la colección que se ubican dentro de los límites de cada intervalo; si se realiza la suma de las frecuencias en los intervalos se obtiene el número total de datos de la colección.

Cálculo de las frecuencias acumulada y relativa

No cabe duda que el análisis descriptivo de datos agrupados es interesante, ya que ofrece la oportunidad de detallar cada uno de los parámetros estadísticos. Para iniciar se pueden calcular algunas variaciones sobre la frecuencia, como son:

- **Frecuencia acumulada** (*fa*), la cual consiste en acumular la frecuencia de los intervalos, de manera que se aprecie la cantidad de datos que se van acumulando conforme se recorren los intervalos de la colección.
- **Frecuencia relativa** (*fr*), la cual indica la proporción porcentual sobre el total del número de datos con los que cuenta cada intervalo.
- **Frecuencia acumulada relativa** (*far*), que representa la proporción porcentual de datos que se va acumulando conforme se recorren los intervalos.

Para ejemplificar lo anterior revisa con detalle el siguiente problema resuelto.

Problema resuelto

Teniendo en cuenta los datos que se muestran en la siguiente tabla de frecuencias, procede a calcular la frecuencia acumulada (*fa*), la frecuencia relativa (*fr*) y la frecuencia acumulada relativa (*far*) de cada intervalo.

Tabla 1.4

Intervalos nominales			*f*
LI		LS	
5	-	9	3
10	-	14	7
15	-	19	14
20	-	24	4
25	-	29	2
$\sum f$	=	n =	30

Alerta

Los datos agrupados se distinguen por estar catalogados en categorías o intervalos.

Estadísticos e-Books & Papers

Solución

A partir de la columna de frecuencias se determinan los valores de las frecuencias solicitadas en cada intervalo.

Tabla 1.5

Intervalos nominales LI		LS	*f*	*fa*	*fr*	*far*
5	–	9	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	4	28	$\frac{4}{30} = 13.33\%$	93.33%
25	–	29	2	30	$\frac{2}{30} = 6.67\%$	100.00%
$\sum f$	=	n =	30		100.00%	

La amplitud de clase

Alerta

La amplitud de clase indica el número de elementos entre los límites de los intervalos.

Por otro lado, existen algunos parámetros importantes por determinar dentro de las tablas de frecuencias, como lo es la **amplitud de clase** (*c*), la cual indica qué tan ancho o amplio es un intervalo, o sea, cuántos valores componen el intervalo, donde la amplitud de clase se determina por la diferencia entre el límite inferior de un intervalo "A" con respecto al límite inferior de un intervalo "B" o en su caso entre los límites superiores.

$$c = LI_B - LI_A \quad \text{o} \quad c = LS_B - LS_A$$

Problema resuelto

Teniendo en cuenta los datos que se muestran en la siguiente tabla de frecuencias, procede a calcular la amplitud de clase.

Tabla 1.6

Intervalos nominales LI		LS	*f*	*fa*	*fr*	*far*
5	–	9	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	4	28	$\frac{4}{30} = 13.33\%$	93.33%

461
Estadísticos e-Books & Papers

Problema resuelto *(continuación)*

Tabla 1.6

Intervalos nominales LI	–	LS	*f*	*fa*	*fr*	*far*
25	–	29	2	30	$\frac{2}{30} = 6.67\%$	100.00%
$\sum f = n =$			30		100.00%	

Solución

En el caso particular del ejemplo la amplitud de clase es 5, para comprobarlo considérense los valores de los límites inferiores del segundo y tercer intervalos, por lo que,

$$c = 15 - 10 = 5$$

Para verificar tómense en cuenta los valores de los límites superiores del cuarto y quinto intervalos:

$$c = 29 - 24 = 5$$

Las marcas de clase

Otro parámetro primordial dentro de las tablas de frecuencias son las denominadas marcas de clase (MC), las cuales son los valores que representan los puntos medios de cada intervalo, donde su valor se determina calculando el promedio de los límites superior e inferior de cada intervalo:

$$MC = \frac{LI+LS}{2}$$

Para ejemplificar lo anterior se muestra el siguiente problema.

Problema resuelto

Teniendo en cuenta los datos que se muestran en la siguiente tabla de frecuencias, procede a calcular las marcas de clase de cada intervalo.

Tabla 1.7

Intervalos nominales LI	–	LS	*f*	*fa*	*fr*	*far*
5	–	9	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	4	28	$\frac{4}{30} = 13.33\%$	93.33%
25	–	29	2	30	$\frac{2}{30} = 6.67\%$	100.00%
$\sum f = n =$			30		100.00%	

Estadísticos e-Books & Papers

Alerta

Las marcas de clase son el punto medio de cada intervalo.

Solución

En el caso particular del ejemplo, los valores de las marcas de clase quedan expuestos en la columna identificada como *MC*.

Tabla 1.8

Intervalos nominales			*f*	*MC*	*fr*	*far*
LI		LS				
5	–	9	3	7	10.00%	10.00%
10	–	14	7	12	23.33%	33.33%
15	–	19	14	17	46.67%	80.00%
20	–	24	4	22	13.33%	93.33%
25	–	29	2	27	6.67%	100.00%
$\sum f$	=	n =	30		100.00%	

Los intervalos de clase reales

Debe notarse que entre los límites superiores e inferiores de los intervalos nominales no se muestra una continuidad, o sea que "existen espacios entre ellos", lo que indica que los datos con valores que estén dentro de estos segmentos no estarían catalogados; para evitar lo anterior se requiere calcular los llamados límites reales, los cuales permiten tener una continuidad en el recorrido de los intervalos, evitando así la indefinición.

Los límites reales se calculan promediando el límite nominal superior de un intervalo "A" con el límite nominal inferior de un intervalo "B", donde los valores del límite inferior real del primer intervalo de clase, así como el límite superior del último intervalo de clase, se pueden determinar restando y sumando la amplitud de clase a los límites reales de los intervalos de clase inmediatos posterior y anterior a los mismos.

Para aclarar lo anterior se expone el siguiente problema resuelto.

Problema resuelto

Teniendo en cuenta los datos que se muestran en la siguiente tabla de frecuencias, procede a calcular los valores de los límites reales de cada intervalo.

Tabla 1.9

Intervalos nominales			*f*	*MC*	*fr*	*far*
LI		LS				
5	–	9	3	7	10.00%	10.00%
10	–	14	7	12	23.33%	33.33%
15	–	19	14	17	46.67%	80.00%
20	–	24	4	22	13.33%	93.33%
25	–	29	2	27	6.67%	100.00%
$\sum f$	=	n =	30		100.00%	

Estadísticos e-Books & Papers

Solución

En el caso particular del ejemplo, los valores de los límites de clase reales se muestran en la columna correspondiente.

Tabla 1.10

Intervalos nominales			Intervalos reales			*F*	*fa*	*fr*	*far*
LI		LS	LIR		LSR				
5	–	9	4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$	93.33%
25	–	29	24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$	100.00%
						$\sum f = n = 30$		100.00%	

1.7 Gráficas descriptivas: histograma, polígono de frecuencias y ojiva

La estadística descriptiva se distingue por sus elementos gráficos, siendo los principales:

a) **Histograma o diagrama de barras**, el cual permite representar la magnitud de la frecuencia de cada uno de los intervalos mediante rectángulos, barras o prismas cuyo largo es el valor de la frecuencia, mientras el ancho es la amplitud de clase delimitado por los límites reales de cada intervalo.

b) **Polígono de frecuencias**, es la línea discontinua que une a las marcas de clase cada intervalo y que permite visualizar la forma en que se han distribuido los datos en correspondencia al recorrido de los intervalos de clase; el polígono de frecuencias da origen a la curva de distribución de frecuencias.

c) **Ojiva**, es una línea discontinua que representa los valores de la frecuencia acumulada relativa (*far*) en razón del recorrido de los intervalos reales de clase de 0% a 100%, expone cómo se acumulan de manera proporcional los datos de una colección, donde los cambios de pendiente representan los cambios en las proporciones acumuladas.

Alerta

Las gráficas descriptivas permiten concentrar información para interpretarla con más facilidad.

Histograma

Para construir un histograma se utiliza un plano X-Y, en cuyas abscisas se representan los límites reales de clase mientras que en las ordenadas se representan las frecuencias. Para desarrollarlo basta con marcar los límites reales de cada intervalo levantando sobre los mismos un rectángulo con una altura que represente el valor de la frecuencia en dicho intervalo.

Para ejemplificar lo anterior considérese el siguiente problema resuelto.

Estadísticos e-Books & Papers

Problema resuelto

Teniendo en cuenta los datos de la siguiente tabla de frecuencias, procede a construir el histograma correspondiente.

Tabla 1.11

Intervalos nominales			Intervalos reales			*f*	*fa*	*fr*	*far*
LI		LS	LIR		LSR				
5	–	9	4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$	93.33%
25	–	29	24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$	100.00%
						$\sum f = n = 30$		100.00%	

Solución

En el caso particular del ejemplo se ubican los valores de los límites reales en el eje de las x, procediendo a levantar las barras de acuerdo con su correspondiente frecuencia.

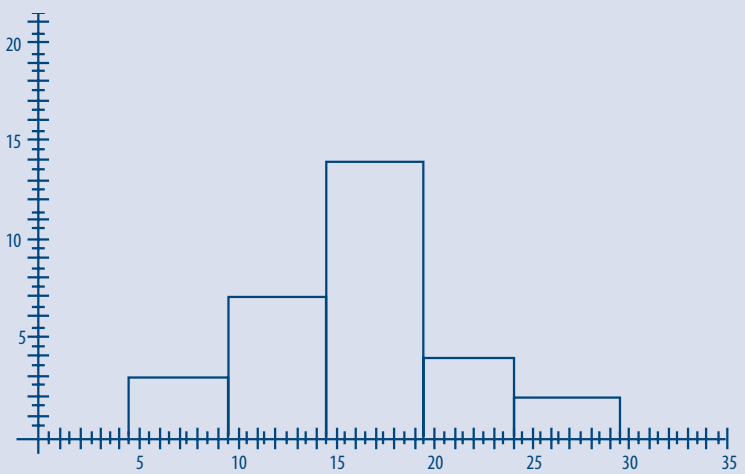


Figura 1.1

Alerta

La construcción del histograma se fundamenta en los límites reales de clase y las frecuencias.

Polígono de frecuencias

El trazo del polígono de frecuencias se fundamenta en el histograma, ya que basta con ubicar las marcas de clase de cada intervalo, representado por el centro de cada barra, ubicando este en la parte superior de cada una, para que posteriormente se unan con una línea discontinua.

A continuación se expone el siguiente problema resuelto para mostrar lo anterior.

465
Estadísticos e-Books & Papers

Problema resuelto

Considerando el polígono de frecuencias que se muestra, procede a ubicar las marcas de clase dentro de cada barra de manera que permita el trazo del polígono de frecuencias.

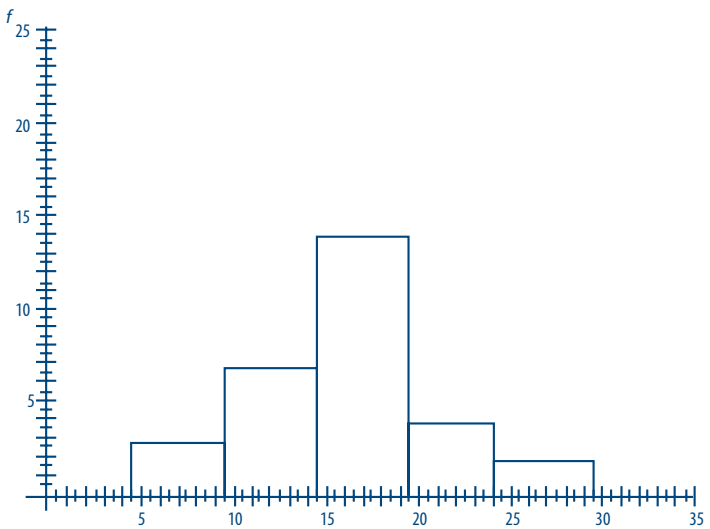


Figura 1.2

Solución

Para ubicar geométricamente las marcas de clase en cada barra se trazan las diagonales en cada rectángulo de manera que se ubiquen los centros de cada una.

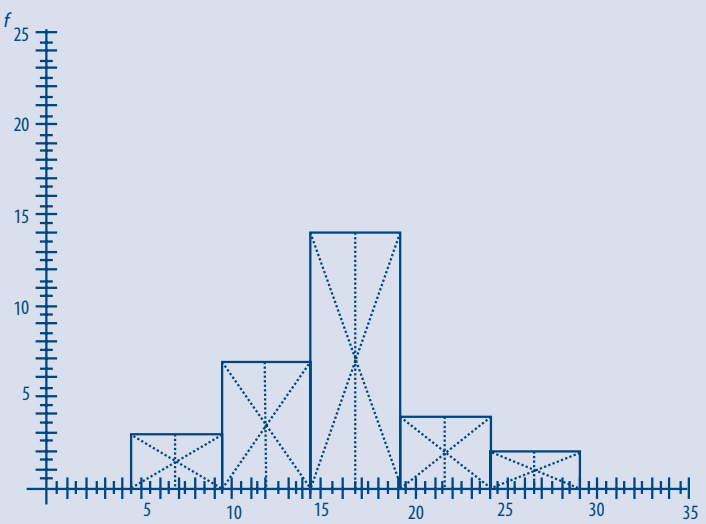


Figura 1.3

Estadísticos e-Books & Papers

Solución *(continuación)*

Uniendo las marcas de clase:

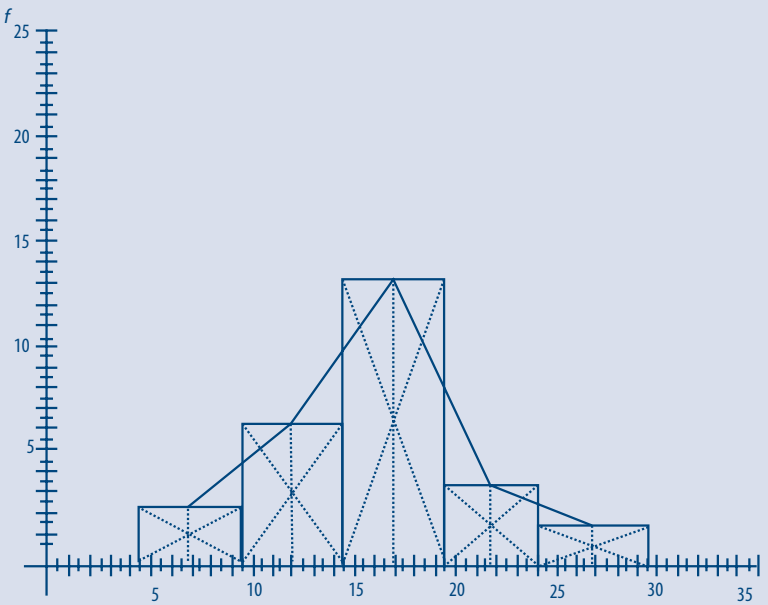


Figura 1.4

Sin embargo, nótese que parece que la línea se encuentra suspendida sobre el histograma, además de que la mitad del primer y último intervalos no se encuentran cubiertas por el polígono, por lo que se deben ubicar unas marcas de clase ficticias en los extremos a efecto de "aterrizar" al polígono procediendo a restar y sumar la amplitud de clase a la primera y última marcas de clase correspondientemente.

Alerta

El polígono de frecuencias permite establecer el perfil de la curva de distribución de frecuencias.

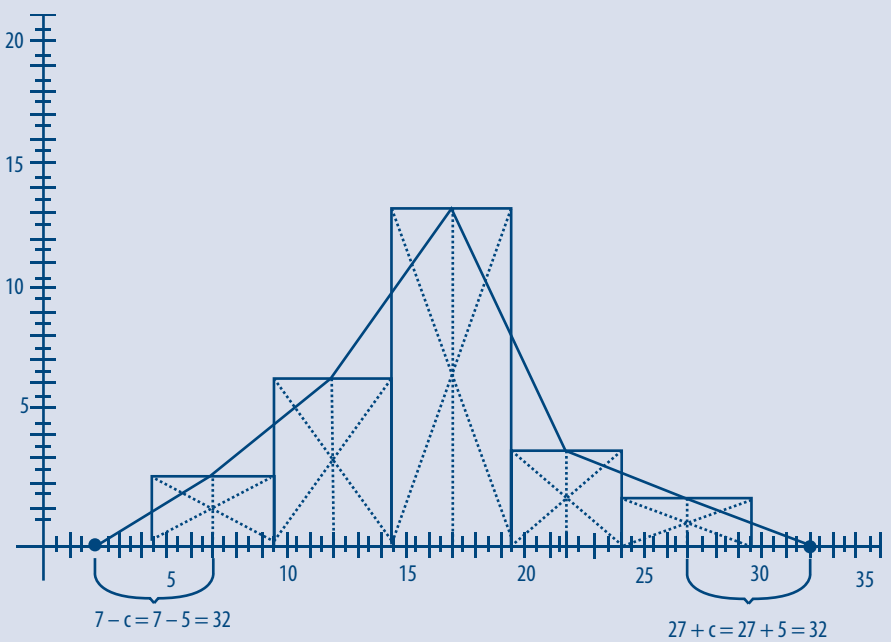


Figura 1.5

Estadísticos e-Books & Papers

La curva de distribución de frecuencias

La importancia del polígono de frecuencias es que expone de manera gráfica cómo se distribuyen los datos de la colección a través de los intervalos de confianza; en realidad, si el trazo del polígono de frecuencias se suaviza da lugar a la llamada curva de distribución de frecuencias. De hecho se cuenta con modelos bien definidos que exponen las diferentes formas en que se distribuyen los datos y que determinan la ubicación de los valores de las medidas de tendencia central, tal como se expone a través de los siguientes casos.

I. **Curva de distribución simétrica.** En este tipo de distribución el intervalo central cuenta con la mayor frecuencia (mayor cantidad de datos), existiendo las mismas frecuencias tanto por arriba como por debajo de este intervalo, además de que la media, la mediana y la moda se concentran al centro de este intervalo compartiendo el mismo valor.

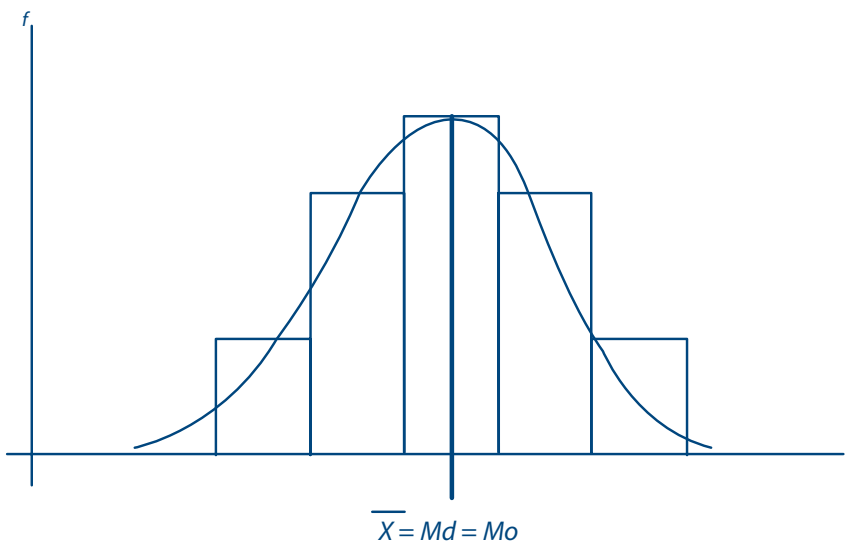


Figura 1.6

II. **Curva de distribución asimétrica a la derecha o con sesgo positivo.** En este tipo de distribución la mayoría de los datos se concentran en los primeros intervalos, donde se ubican los valores bajos de la colección, mientras que el área de asimetría se ubica en los intervalos donde se concentran los valores altos de la colección (hacia la derecha). En este esquema el orden de las medidas de tendencia central es el siguiente: moda, mediana y media.

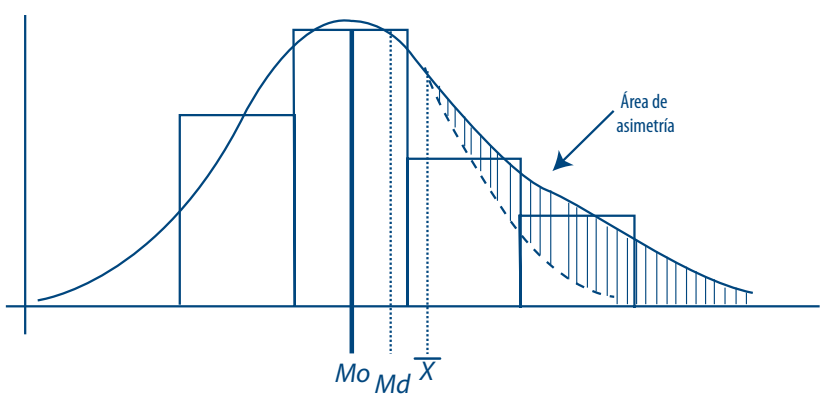


Figura 1.7

Estadísticos e-Books & Papers

III. **Curva de distribución asimétrica a la izquierda o sesgo negativo.** En este tipo de distribución la mayoría de los datos se concentran en los últimos intervalos donde se ubican los valores altos de la colección, mientras que el área de asimetría se ubica en los intervalos donde se concentran los valores bajos de la colección (a la izquierda). En este esquema el orden de las medidas de tendencia central es el siguiente: media, mediana y moda.

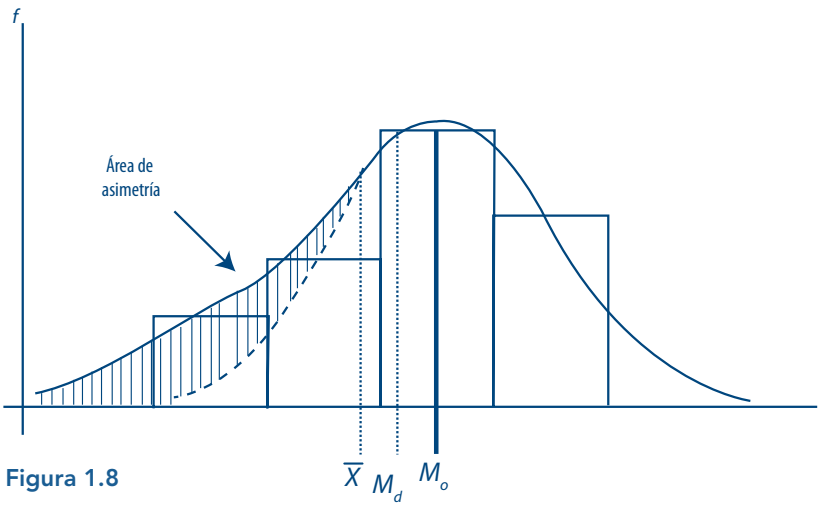
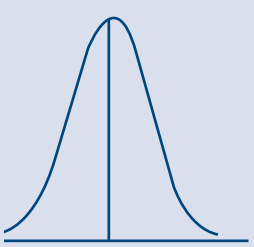
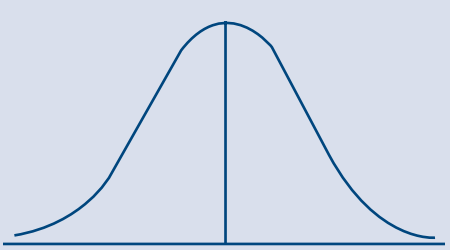
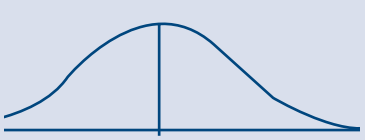


Figura 1.8

Además, el polígono de frecuencias también permite establecer *a priori* qué tan dispersos se encuentran los datos con respecto a la media, ya que con base en esto la curva de distribución tomará ciertas configuraciones, las cuales se denominan de apuntamiento y se exponen a continuación:

Tipo de apuntamiento	Interpretación
Leptocúrtica 	Los valores de los datos están cercanos a la media, habiendo desviaciones mínimas y, en consecuencia, se presenta una baja dispersión, haciendo que la curva de distribución sea esbelta y erguida.
Mesocúrtica 	Los valores de los datos se disponen de manera moderada alrededor de la media; están dispersos de manera moderada.
Platocúrtica 	Los valores de los datos se encuentran bastante alejados de la media, por lo que se presentan grandes desviaciones, por lo que la curva de distribución es amplia y de baja altura.

469
Estadísticos e-Books & Papers

La ojiva

Es la gráfica que permite establecer los valores acumulados de frecuencia de manera porcentual, o sea, se fundamenta en la ***far***. La ojiva se compone de segmentos de recta correspondientes a cada intervalo de clase, donde la pendiente de estos segmentos indica el aumento o la disminución de la frecuencia acumulada relativa.

Considerando los valores de los límites reales así como de la ***far***, la ojiva se construye marcando los valores de la *far* sobre los límites superiores de cada intervalo, tal como se muestra a continuación.

Problema resuelto

Considerando la siguiente tabla de frecuencias, procede a construir la ojiva correspondiente.

Tabla 1.12

Intervalos nominales			Intervalos reales			*f*	*fa*	*fr*	*far*
LI		LS	LIR		LSR				
5	–	9	4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$	93.33%
25	–	29	24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$	100.00%
						$\sum f = n = 30$		100.00%	

Solución

Como se mencionó anteriormente, se marcan los límites de clase de los intervalos de clase reales, procediendo a ubicar las ***far*** correspondientes tal como se muestra.

Nótese que el cambio en la pendiente del segmento del tercer intervalo se debe al marcado cambio en el valor de la ***far***, ya que pasa de 33.33% a 80%.

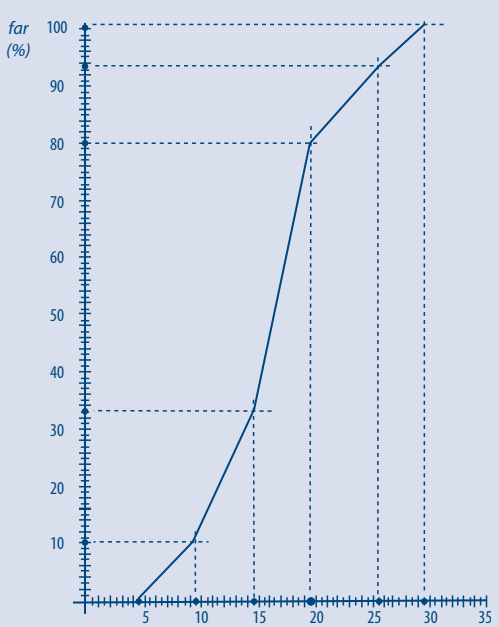


Figura 1.9

Alerta

La ojiva permite establecer la *far* de valores de interés dentro del rango de la colección.

Estadísticos e-Books & Papers

1.8 Medidas de tendencia central para datos agrupados

Alerta

El cálculo de las medidas de tendencia central se fundamenta en los límites de clase real, la frecuencia, la frecuencia acumulada y la amplitud de clase.

El cálculo de los valores de las medidas de tendencia para datos agrupados puede causar algo de dificultad en el entendido de que estos se encuentran implícitos dentro de los límites de los intervalos que los contienen, ya que tan solo se cuenta con el número de datos que se han catalogado en cada intervalo, o sea las frecuencias; sin embargo, con base en las mismas es posible determinar sus valores.

De hecho, los procesos de cálculo consisten en determinar la proporción de la amplitud de clase que separa el valor de la medida de centralización con respecto al límite inferior real del intervalo donde se encuentra; la distancia se determina mediante el cálculo de una razón compuesta por las frecuencias reales y acumuladas del intervalo.

Cálculo de la media para datos agrupados

Para determinar el valor de la media se aplicará el principio: "el promedio de los promedios es el promedio", de manera que al determinar el promedio de las veces en que las marcas de clase se repiten dentro de los intervalos se obtendrá la media:

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i \; MC_i}{n}$$

donde:

f_i = Valor de la frecuencia en el intervalo i (i = 1, 2, 3… n).

MC_i = Valor de la marca de clase del intervalo i.

N = Número de datos de la colección.

Alerta

La media de datos agrupados se fundamenta en el promedio de las marcas de clase.

Problema resuelto

Considerando la siguiente tabla de frecuencias, procede a determinar el valor de la media.

Tabla 1.13

Intervalos nominales			Intervalos reales			*f*	*fa*	*fr*	*far*
LI		LS	LIR		LSR				
5	–	9	4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$	93.33%
25	–	29	24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$	100.00%
						$\sum f = n = 30$		100.00%	

Solución

Por tanto, de acuerdo con el ejemplo en desarrollo, el valor de la media es:

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i \; MC_i}{n} = \frac{485}{30} = 16.67$$

Estadísticos e-Books & Papers

Cálculo de la mediana para datos agrupados

El cálculo de la mediana se fundamenta en el límite inferior real y las frecuencias donde se sitúa el valor de la frecuencia acumulada relativa del 50%, debido a que el valor de la mediana de una colección de datos se ubica al centro de la misma.

$$M_d = LIR_{M_d} + \frac{\frac{n}{2} - fa_{antM_d}}{f_{M_d}} \cdot c$$

donde

LIR_{Md} = Límite inferior real de la clase mediana.

fa_{antMd} = Frecuencia acumulada de la clase anterior a la clase mediana.

f_{Md} = Frecuencia de la clase mediana.

c = Amplitud de clase de la clase mediana.

n = Número total de datos de la colección.

Por tanto, para ejemplificar se presenta el siguiente problema.

Problema resuelto

Considerando la siguiente tabla de frecuencias, procede a determinar el valor de la mediana.

Tabla 1.14

Intervalos nominales			Intervalos reales			*f*	*fa*	*fr*	*far*
LI		LS	LIR		LSR				
5	–	9	4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$	93.33%
25	–	29	24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$	100.00%
						$\sum f = n = 30$		100.00%	

Solución

Considerando los datos de la tabla de frecuencias nos damos cuenta de que, con base en los valores de la ***far,*** 50% de la colección se encuentra ubicado en el tercer intervalo, por lo que aplicando la fórmula para el cálculo de la mediana:

$$M_d = LIR_{M_d} = \frac{\frac{n}{2} - f_{antM_d}}{f_{M_d}} \cdot c = 14.5 + \frac{\frac{30}{2} - 10}{14} (5) = 16.28$$

Alerta

La mediana es el valor que representa una *far* igual a 50%.

Cálculo de la moda para datos agrupados

En relación con la moda (M_o), se fundamenta en ubicar el intervalo o intervalos que tengan la frecuencia mayor entre todas las frecuencias, en consideración a que en esa o esas clases es donde se puede repetir uno o varios valores. El cálculo de la moda se logra a través de la siguiente fórmula:

Estadísticos e-Books & Papers

$$M_o = LIR_{M_o} + \frac{(f_{M_o} - f_1)}{(f_{M_o} - f_1) + (f_{M_o} - f_2)} \cdot c = LIR_{M_o} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot c$$

donde

LIR_{Mo} = Límite inferior real de la clase modal.
f_{Mo} = Frecuencia de la clase modal.
f_1 = Frecuencia anterior a la clase modal.
f_2 = Frecuencia posterior a la clase modal.
c = Amplitud de clase de la clase modal.

Problema resuelto

Considerando la siguiente tabla de frecuencias, procede a determinar el valor de la moda.

Tabla 1.15

Intervalos nominales			Intervalos reales			f	fa	fr	far
LI		LS	LIR		LSR				
5	–	9	4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$	10.00%
10	–	14	9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$	33.33%
15	–	19	14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$	80.00%
20	–	24	19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$	93.33%
25	–	29	24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$	100.00%
						$\sum f = n = 30$		100.00%	

Solución

En atención a la tabla de frecuencias del ejemplo en desarrollo se determina que la colección de datos es modal, considerando que la frecuencia con mayor valor se ubica en el tercer intervalo, por lo que, sustituyendo valores en la fórmula, se tiene que:

$$M_o = LIR_{M_o} + \frac{(f_{Mo} - f_1)}{(f_{M_o} - f_1) + (f_{M_o} - f_2)} \cdot c = 14.5 + \frac{(14 - 7)}{(14 - 7) + (14 - 4)} \cdot (5) = 16.56$$

Alerta

La moda se ubica en el intervalo o intervalos con la frecuencia mayor.

1.9 Determinación de la ubicación y el valor de la moda de manera gráfica mediante el histograma y el polígono de frecuencias

Mediante el histograma y el polígono de frecuencias se puede determinar la ubicación y el valor aproximado de la moda. En este caso particular se debe ubicar el intervalo de la clase mediana de manera que se trace una diagonal que una los vértices superiores derechos de las barras del intervalo inmediato anterior con el de la clase modal, procediendo de igual manera con los vértices superiores izquierdos del intervalo inmediato posterior y el de la clase modal, de manera que la línea perpendicular a la intersección de las diagonales permitirá ubicar el valor, tal como se muestra a continuación en el siguiente problema resuelto.

Estadísticos e-Books & Papers

Problema resuelto

Considerando el histograma y polígono de frecuencias que se muestran, procede a determinar el valor de la moda de manera gráfica.

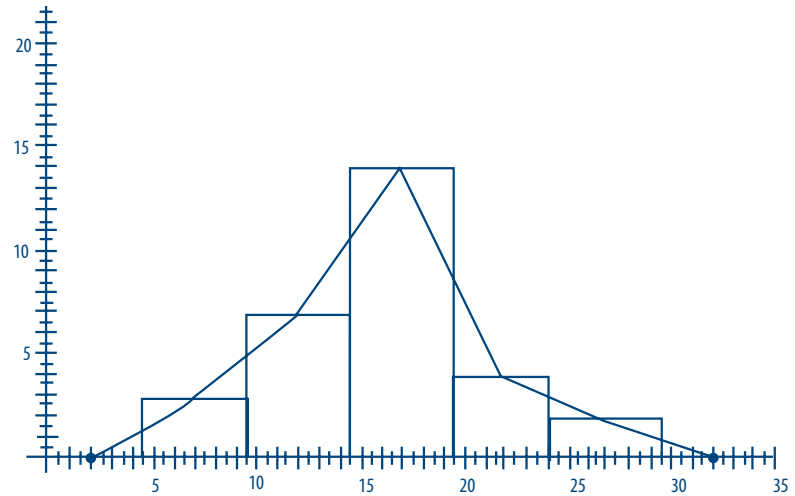


Figura 1.10

Solución

Desarrollando el método descrito se determina el valor de la moda dentro del intervalo con la mayor frecuencia.

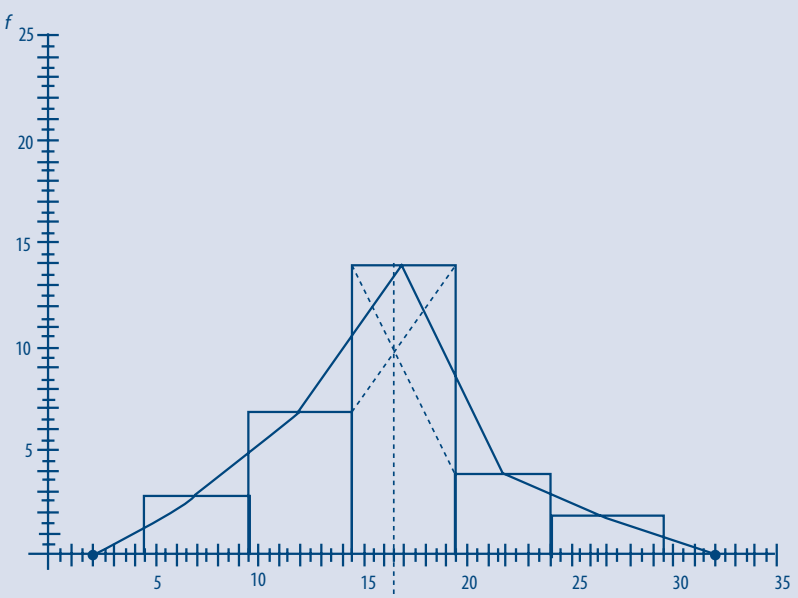


Figura 1.11

Alerta

El método gráfico permite verificar el resultado del cálculo de la moda.

Estadísticos e-Books & Papers

1.10 Determinación de la ubicación y el valor de la mediana de manera gráfica mediante la ojiva

Para determinar el valor de la mediana por medio de la ojiva es necesario trazar lo que algunos denominan la ojiva descendente, donde los valores de la ***far*** para la misma están definidos por ***1-far***. El valor de la mediana corresponde a la abscisa de la intersección de las dos ojivas, tal como se demuestra en el siguiente problema.

Problema resuelto

Considerando la información que se proporciona en la siguiente tabla de frecuencias, procede a determinar el valor de la mediana de manera gráfica.

Tabla 1.16

Intervalos reales			*f*	*fa*	*fr*
LIR		LSR			
4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$
9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$
14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$
19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$
24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$
			$\sum f = n = 30$		100.00%

Solución

Desarrollando el método descrito se determina el valor de la mediana.

Tabla 1.17

Intervalos reales			*f*	*fa*	*fr*	*far I*	*far II*
LIR		LSR					
4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$	10.00%	1 – 0.10 = 90.00%
9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$	33.33%	1 – 0.3333 = 66.67%
14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$	80.00%	1 – 0.8 = 20.00%
19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$	93.33%	1 – 0.9333 = 6.67%
24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$	100.00%	1 – 1 = 0.00%
			$\sum f = n = 30$		100.00%		

Estadísticos e-Books & Papers

Solución *(continuación)*

Trazando las ojivas de acuerdo con los límites de clase correspondientes:

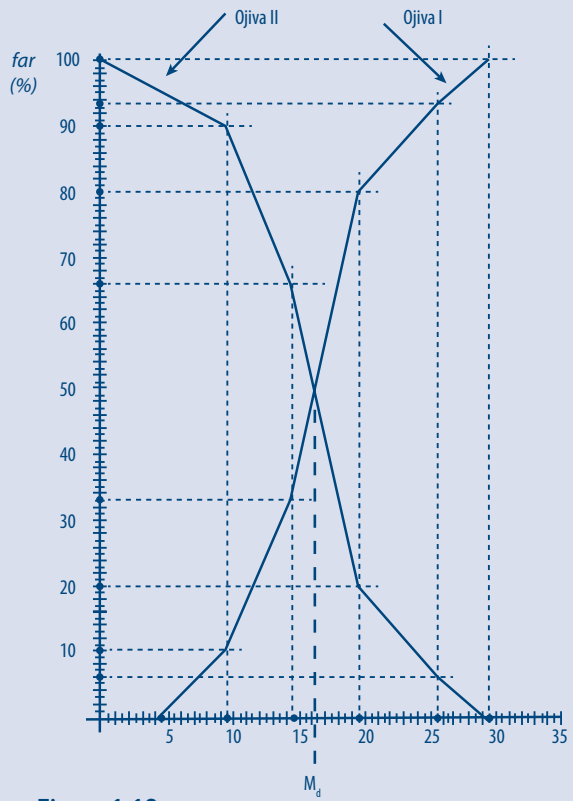


Figura 1.12

Alerta

Debe observarse la construcción de las ojivas a efecto de lograr la correcta ubicación del valor de la mediana.

1.11 Cálculo de cuantiles para datos agrupados

El cálculo de los cuantiles para colecciones de datos agrupados se fundamenta en la determinación del valor de los percentiles, por lo que para determinar el valor de un cuartil o decil en particular se deberá considerar su equivalencia en percentiles. El valor de cualquier percentil se puede calcular por medio de la fórmula que da origen a la fórmula para la determinación del valor de la mediana.

$$P_p = LIR_P + \frac{\left(\frac{P}{100} \cdot n\right) - fa_{antP}}{f_P} \cdot c$$

donde

LIR_P = Límite inferior real de la clase que contiene al percentil P.
fa_{antP} = Frecuencia acumulada de la clase anterior a la clase percentil.
f_P = Frecuencia de la clase percentil.
c = Amplitud de clase de la clase percentil, recordando que se trata de una constante.
n = Número total de datos de la colección.

Para ejemplificar la aplicación de la fórmula anterior considérese el siguiente problema.

Estadísticos e-Books & Papers

Problema resuelto

Considerando la siguiente tabla de frecuencias, procede a determinar los valores del tercer cuartil y del segundo decil.

Tabla 1.18 Frecuencias

Intervalos reales			*f*	*fa*	*fr*
LIR		LSR			
4.5	–	9.5	3	3	$\frac{3}{30} = 10.00\%$
9.5	–	14.5	7	10	$\frac{7}{30} = 23.33\%$
14.5	–	19.5	14	24	$\frac{14}{30} = 46.67\%$
19.5	–	24.5	4	28	$\frac{4}{30} = 13.33\%$
24.5	–	29.5	2	30	$\frac{2}{30} = 6.67\%$
			$\sum f = n = 30$		100.00%

Solución

Considerando los valores de la ***far*** se puede determinar que el tercer cuartil (75%) se ubica en el tercer intervalo, mientras que el segundo decil (20%) se ubica en el segundo intervalo, por lo que, sustituyendo en la fórmula:

$$Q_3 = P_{75} = LIR_P + \frac{\left(\frac{P}{100} \cdot n\right) - fa_{antP}}{f_p} \cdot c = 14.5 + \frac{\left(\frac{75}{100} \cdot 30\right) - 10}{14}(5) = 18.96$$

$$D_2 = P_{20} = LIR_P + \frac{\left(\frac{P}{100} \cdot n\right) - fa_{antP}}{f_p} \cdot c = 9.5 + \frac{\left(\frac{20}{100} \cdot 30\right) - 3}{7}(5) = 11.64$$

Alerta

La fórmula de los cuantiles para datos agrupados ofrece el valor del percentil de interés.

1.12 Momentos estadísticos

Considerando que la configuración de la curva de distribución de frecuencias ofrece la disposición de los datos en razón del apuntamiento y simetría, mismo que depende de los valores de las frecuencias y de la dispersión, es posible establecer la caracterización de la curva sin necesidad de desarrollar los elementos gráficos; esta opción la ofrecen los llamados momentos estadísticos con respecto a la media.

Los momentos estadísticos se fundamentan en las diferencias entre las marcas de clase y la media, considerando que las marcas de clase son los valores representativos de cada intervalo, como lo es también la media pero de toda la colección. Estas diferencias bajo condiciones promedio permitirán establecer las características de la curva tal como se expone a continuación.

El cálculo fundamental es la diferencia de las marcas de clase con respecto a la medía:

$$Y = MC - \overline{X}$$

Estadísticos e-Books & Papers

De manera que los momentos estadísticos (*ME*) se definen e interpretan como sigue:

a) Momento estadístico de primer grado, el cual expone que las diferencias entre las marcas de clase con respecto a la media deben ser cero o próximas a él, ya que los valores de las desviaciones por arriba y por debajo de la media deben ser iguales.

$$ME_1 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i}{n} = 0$$

donde

f_i = Frecuencia del intervalo i (i = 1, 2, 3,… n).

Y_i = Diferencia de marca de clase del intervalo i con respecto a la media.

n = Número total de datos de la colección.

Alerta

El momento de primer grado permite corroborar que el promedio de las diferencias con respecto a la media es cero.

b) Momento de segundo grado, que expone el promedio de las desviaciones al cuadrado de las marcas de clase con respecto a la media, lo que significa que se está calculando la varianza, por lo que es posible determinar el valor de la desviación estándar (S).

$$ME_4 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^2}{n} = S^4$$

donde

f_i = Frecuencia del intervalo i (i = 1, 2, 3,… n).

Y_i = Diferencia de marca de clase del intervalo i con respecto a la media elevada al cuadrado.

n = Número total de datos de la colección.

Alerta

El momento de segundo grado permite determinar el valor de la desviación estándar con base en la suma de los cuadrados de las variaciones entre las marcas de clase y la media.

c) Momento estadístico de tercer grado, en el que las desviaciones de las marcas de clase con respecto a la media se elevan al cubo promediándolas con respecto al número de elementos de la colección, lo que permite establecer el tipo de sesgo que guarda la curva de distribución a través del coeficiente de asimetría (K_3).

$$ME_3 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^3}{n}$$

donde

f_i = Frecuencia del intervalo i (i = 1, 2, 3,… n).

Y_i = Diferencia de marca de clase del intervalo i con respecto a la media elevada al cubo.

n = Número total de datos de la colección.

Con base en el momento de tercer grado se calcula el coeficiente de asimetría:

$$k_3 = \frac{ME_3}{S^3}$$

donde

Si $k_3 > 0$, la curva de distribución tiene una asimetría derecha o sesgo positivo.

Si $k_3 < 0$, la curva de distribución tiene una asimetría izquierda o sesgo negativo.

Si $k_3 = 0$, la curva de distribución es simétrica.

Alerta

El momento de tercer grado determina el sesgo de la curva de la distribución de frecuencias.

d) Momento de cuarto grado, en el que las desviaciones de las marcas de clase con respecto a la media se elevan a la cuarta potencia, promediándose con respecto al número de elementos de la colección, lo que permite establecer el tipo de apuntamiento de la curva de distribución de frecuencias a través del coeficiente de apuntamiento de curtosis.

$$ME_4 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^2}{n} = S^4$$

Alerta

El momento de cuarto grado determina el apuntamiento de la curva de la distribución de frecuencias.

Estadísticos e-Books & Papers

donde

f_i = Frecuencia del intervalo i (i = 1, 2, 3,… n).

Y_i = Diferencia de marca de clase del intervalo i con respecto a la media elevada a la cuarta potencia.

n = Número total de datos de la colección.

Con base en el momento de cuarto grado se calcula el coeficiente de apuntamiento o de curtosis:

$$k_4 = \frac{ME_4}{S^4}$$

De manera que

Si $k_4 - 3 > 0$, la cuva es leptocúrtica.

Si $k_4 - 3 = 0$, la curva es mesocúrtica.

Si $k_4 - 3 < 0$, la curva es platocúrtica.

Problema resuelto

Considerando la siguiente tabla de frecuencias, procede a determinar la caracterización de la curva de distribución de frecuencias por medio del cálculo de momentos estadísticos.

Tabla 1.19

Intervalos nominales		Intervalos reales		MC	f	fa	far	f * MC	$Y = MC - \bar{X}$
5	9	4.5	9.5	7	3	3	10.00%	21.00	–9.17
10	14	9.5	14.5	12	7	10	33.33%	84.00	–4.17
15	19	14.5	19.5	17	14	24	80.00%	238.00	0.83
20	24	19.5	24.5	22	4	28	93.33%	88.00	5.83
25	29	24.5	29.5	27	2	30	100.00%	54.00	10.83
				$\sum f = n =$	**30**		$\sum =$	**485**	

Solución

Tabla 1.20

Intervalos nominales		Intervalos reales		f	$Y = MC - \bar{X}$	f*Y	$f*Y^2$	$f*Y^3$	$f*Y^4$
5	9	4.5	9.5	3	−9.17	−27.50	252.08	−2310.76	21 182.00
10	14	9.5	14.5	7	−4.17	−29.17	121.53	−506.37	2 109.86
15	19	14.5	19.5	14	0.83	11.67	9.72	8.10	6.75
20	24	19.5	24.5	4	5.83	23.33	136.11	793.98	4 631.56
25	29	24.5	29.5	2	10.83	21.67	234.72	2542.82	27 547.26
			$\sum f = n =$	**30**	$\sum =$	**0.00**	**754.17**	**527.78**	**55 477.43**

Donde las características de la curva de distribución de frecuencias a través de los momentos se exponen a continuación.

$$ME_1 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i}{n} = \frac{0}{30} = 0$$

$$ME_2 = S_2 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^2}{n} = \frac{754.17}{30} = 25.14$$

$$S = \sqrt{25.14} = 5.014$$

$$ME_3 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^3}{n} = \frac{527.78}{30} = 17.59$$

$$k_3 = \frac{ME_3}{S^3} = \frac{17.59}{5.014^3} = 0.134$$

Estadísticos e-Books & Papers

Solución *(continuación)*

Si $k_3 > 0$, la curva de distribución tiene una asimetría derecha o sesgo positivo.

$$ME_4 = \frac{\sum_{i=1}^{n} f_i \cdot Y_i^4}{n} = \frac{55477.43}{30} = 1849.25$$

$$k_4 = \frac{ME_4}{S^4} = \frac{1849.25}{5.014^4} = 2.92$$

Si $k_4 - 3 = -0.08$, la curva es ligeramente platocúrtica.

1.13 Determinación del número de intervalos de clase

La determinación del número de intervalos para realizar un estudio de estadística descriptiva puede generar puntos de vista encontrados y ser motivo de discusión. Sin embargo, es posible determinarlo de manera simple si se observan el cumplimiento del proceso de cálculo y el uso del buen criterio.

Debe tenerse presente que un número muy amplio de intervalos propiciaría la obtención de demasiados valores que no ofrecerían una buena perspectiva de análisis; tampoco favorece al análisis tener un número muy limitado de intervalos.

En esta obra se presentan dos métodos simples, mismos que se exponen a continuación. El primero de ellos es el método empírico, el cual consiste en que una vez que se haya determinado el valor del rango de una colección se buscará un par de números cuyo producto se aproxime al mismo, considerando que el primer número represente el número de intervalos y el segundo la amplitud de clase; para ejemplificar este método considérese el siguiente problema resuelto.

Problema resuelto

Considerando la colección de datos que se muestra, procede a determinar el número de intervalos por el método empírico.

169	128	140	176	161	119	164	154
140	136	154	136	148	144	175	148
136	162	130	141	123	159	151	150
146	139	160	146	136	148	144	152
145	142	156	145	128	150	135	161

Solución

En la colección de datos el valor máximo es 176 mientras que el valor mínimo es 119, por lo que,

$$R = 176 - 119 = 57$$

Revisando las tablas de multiplicar:

Tabla 1.21

Número de intervalos	Amplitud de clase
1	56
56	1
2	28
28	2
4	14
14	4
7	8
8	7

Estadísticos e-Books & Papers

Solución *(continuación)*

Analizando las opciones se considera que las dos últimas representan las mejores; por optar por alguna se elige la primera de ellas, por lo que se calcula la amplitud de clase con el propósito de ajustarla de ser necesario.

$$\textit{Amplitud de clase} = c = \frac{\textit{Rango}}{\textit{No. de intervalos}} = \frac{57}{7} = 8.1 \approx 9$$

Se sube a la siguiente unidad debido a que $7 \times 8 = 56$ que es menor que 57; de manera que $7 \times 9 = 63$ excede el rango de la colección propuesta.

Por ello, los valores de los límites nominales de los intervalos de clase se determinan considerando que el límite inferior del primer intervalo debe empezar en 119, pero quedaría muy justo, por lo que se puede sugerir restar una unidad, de modo que empezaría en 118 y partiendo de este número se le sumará la amplitud de clase para definir los valores de los límites inferiores de clase. Para determinar el valor de los límites superiores de cada intervalo se aplica la siguiente fórmula:

$$LS = (LI + c) - 1$$

La razón de restar una unidad es que el intervalo empieza en el valor del límite inferior, por lo que si $c = 9$,

Tabla 1.22

Límite inferior		Límite superior
118	-	126
127	-	135
136	-	144
145	-	153
154	-	162
163	-	171
172	-	180

El segundo método de cálculo para determinar el número de intervalos es aplicando la fórmula de Sturges:

Problema resuelto

Considerando la colección de datos que se muestra, se procede a determinar el número de intervalos por el método de la fórmula de Sturges.

$$\textit{Número de intervalos} = 1 + 3.3 \cdot \log(n)$$

169	128	140	176	161	119	164	154
140	136	154	136	148	144	175	148
136	162	130	141	123	159	151	150
146	139	160	146	136	148	144	152
145	142	156	145	128	150	135	161

Solución

Considerando que el número de elementos de la colección del ejemplo anterior es 40.

$$\textit{Número de intervalos} = 1 + 3.3 \cdot \log(40) = 6.28 \approx 7$$

Como se aprecia, hay cierta similitud entre los resultados de los métodos expuestos.

481
Estadísticos e-Books & Papers

Solución *(continuación)*

Una vez que se han definido los intervalos se procede a catalogar los datos de la colección en los mismos.

169	128	140	176	161	119	164	154
140	136	154	136	148	144	175	148
136	162	130	141	123	159	151	150
146	139	160	146	136	148	144	152
145	142	156	145	128	150	135	161

Tabla 1.23

Límite inferior		Límite superior		*f*
118	-	126	||	2
127	-	135	||||	4
136	-	144	|||| |||| |	11
145	-	153	|||| |||| |	11
154	-	162	|||| |||	8
163	-	171	||	2
172	-	180	||	2
			$\Sigma =$	40

Una vez completado este paso la tabla de frecuencias está dispuesta para desarrollar un análisis por estadística descriptiva, tal como se ha propuesto a lo largo de esta unidad.

1.14 La media geométrica

Es un valor de tendencia central representativo de una colección de datos cuyos valores guardan estrecha relación unos con otros y observan un orden secuencial bien definido, como lo son la tasa de inflación, la tasa de crecimiento, los factores de devaluación y la tasa de interés, por citar algunos entre los más importantes.

El valor de la media geométrica puede calcularse tanto para datos no agrupados como para datos agrupados.

En el caso de datos no agrupados la media geométrica es la raíz *n*-ésima del producto de los valores del conjunto de datos.

$$\overline{X}_G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \ldots x_n}$$

Considérese el siguiente ejemplo de aplicación.

Problema resuelto

A un jubilado le proponen invertir parte de su pensión mensual en un producto de inversión a cuatro meses que le ofrece pagar 8% el primer mes, 12% el segundo mes, 11% el tercer mes y 8% el cuarto mes. ¿Cuál sería el valor de la tasa promedio que recibiría?

Solución

Si se considera que el fundamento de las matemáticas financieras en el manejo de la tasa de interés a futuro es: $(1+i)^n$, donde i es la tasa de interés y n el número de periodos de aplicación de la tasa, que en este caso es "1", por tanto:

$$\overline{X}_G = \sqrt[4]{1.08 \cdot 1.12 \cdot 1.11 \cdot 1.08} = 1.0974$$

Por lo que restando 1 al resultado la tasa promedio es de 9.74%.

Alerta

Los métodos para la determinación de los intervalos de clase pueden coincidir, considerando la importancia del criterio del analista en la decisión.

Alerta

La media geométrica ofrece un criterio adicional para el análisis estadístico, dependiendo del criterio que se pretenda cubrir.

Estadísticos e-Books & Papers

En el caso de datos agrupados el cálculo del valor de la media geométrica se determina mediante la siguiente fórmula:

$$\ln(\overline{X}_G) = \frac{\sum_{i=1}^{n} f_i \cdot \ln(MC_i)}{n}$$

Considérese el siguiente ejemplo de aplicación.

Problema resuelto

Considerando la siguiente tabla de frecuencias, procede a determinar el valor de la media geométrica.

Tabla 1.24

Intervalos nominales	*f*	MC
2-4	2	3
5-7	5	6
8-10	3	9
Σ	10	

Solución

Se complementan los cálculos necesarios para determinar el valor de la media geométrica.

Tabla 1.25

Intervalos nominales	*f*	MC	ln (MC)	*f* x ln (MC)
2-4	2	3	1.0987	2.1974
5-7	5	6	1.7918	8.9590
8-10	3	9	2.1972	6.5916
Σ	10			17.748

Aplicando la fórmula de la media geométrica para datos agrupados:

$$\ln(\overline{X}_G) = \frac{\sum_{i=1}^{n} f_i \cdot \ln(MC_i)}{n} = \frac{17.748}{10} = 1.7748$$

Obteniendo el antilogaritmo:

$$antiln\ (1.7748) = 5.90$$

483
Estadísticos e-Books & Papers

1.1 Ordena la siguiente colección de datos y calcula el valor de las medidas de tendencia central.

98 23 56 37 78 34 84 16 67

1.2 Ordena la siguiente colección de datos y calcula las medidas de tendencia central.

142 161 150 145 130 184 176 154 149 178
112 90 162 77 198 181 116 100 165 199

1.3 Ordena la siguiente colección de datos y calcula las medidas de tendencia central.

1234 1178 1340 1189 1567
1645 1934 1937 1756 1111
1023 1404 1023 987 1032

1.4 Considerando la colección del problema 1.1, calcula las medidas de dispersión: rango, varianza y desviación estándar.

1.5 Considerando la colección de datos del problema 1.2, calcula las medidas de dispersión: rango, varianza y desviación estándar.

1.6 Considerando la colección del problema 1.3, calcula las medidas de dispersión: rango, varianza y desviación estándar.

1.7 Considerando la colección de datos del problema 1.1, calcula el valor de Q_1, Q_2 y Q_3.

1.8 Considerando la colección de datos del problema 1.2, calcula el valor de los deciles: D_1, D_4, D_6 y D_8.

1.9 Con base en la colección de datos del problema 1.3, calcula y comprueba el valor de la mediana en razón de: $M_d = Q_2 = D_5$.

1.10 Con base en la colección de datos del problema 1.2, calcula los siguientes cuantiles: P_{15}, P_{40} y P_{80}.

1.11 Con base en la colección de datos del problema 1.3, calcula los siguientes cuantiles: Q_1, Q_2, Q_3, P_{32}, P_{54} y P_{95}.

1.12 Considerando la colección de datos del problema 1.1, determina el valor de la mediana de manera directa y verifica su valor con la fórmula de los cuantiles para datos agrupados.

SR **1.13** Una empresa de mantenimiento de equipo industrial hace un recuento de los rodamientos que fueron sustituidos en equipos de bombeo clasificándolos por su diámetro en milímetros. Considerando la información de la siguiente tabla de frecuencias, calcula los intervalos reales de clase y las marcas de clase, así como la amplitud de clase.

Tabla 1.26

Intervalos	*f*
221-230	4
231-240	6
241-250	9
251-260	6
261-270	9
271-280	5
281-290	2
291-300	9
301-310	4

SR **1.14** Una fábrica de herrajes para construcción desarrolla un proceso de estadística descriptiva sobre la longitud de su inventario de tensores para tubería de ½" de diámetro, tal como se expone en la siguiente tabla.

Tabla 1.27

Longitud (cm)	Cantidad
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8

Calcula los intervalos reales y la amplitud de clase, adicionalmente determina las marcas de clase.

1.15 Una empresa importadora de instrumentos de medición y control de procesos químicos clasifica a sus proveedores por el monto de las operaciones de importación anuales, tal como se muestra en la siguiente tabla de frecuencias.

Tabla 1.28

Monto de las operaciones (millones de pesos)	Número de proveedores
De $2 a $ 5	6
De $5 a $ 8	13
De $8 a $11	20
De $11 a $14	10
De $14 a $17	1

Calcula la amplitud de clase y las marcas de clase.

SR **1.16** Una empresa de elementos prefabricados de madera procedió a clasificar los excedentes de bastidores de una pulgada de espesor por su longitud en centímetros. Donde el detalle del inventario se expone a continuación.

25 37 47 60 74
34 38 52 63 66
27 38 49 61 72
46 64 70 44 45
41 53 62 67 42
45 59 71 72 60
50 58 51 56 52
45 49 52 53 57

Como analista de operaciones industriales le ha sido encomendado el desarrollo de un análisis por estadística descriptiva de manera que clasifique el inventario en 5 intervalos, donde el valor del límite inferior nominal del primer intervalo sea el valor mínimo de la colección y el valor del límite superior nominal del quinto intervalo sea el valor máximo de la colección; para que una vez cubierta esta condición se elaboren la tabla de frecuencias, los intervalos de clase reales y las marcas de clase.

SR **1.17** Al término de un proyecto de construcción una empresa de proyectos electromecánicos realiza el levantamiento del inventario de los tramos de cable de 500 MCM considerando su largo en pulgadas. Los responsables del departamento de Ingeniería de Procura (compras) estructuran la muestra que se detalla a continuación.

Estadísticos e-Books & Papers

37	38	38	64	53	59	58	49
74	66	72	45	42	60	52	57
60	63	61	44	67	72	56	53
25	34	27	46	41	45	50	45
65	73	39	70	54	73	71	61
47	52	49	70	62	71	51	52

Elabora la tabla de frecuencias y los intervalos de clase reales, así como las marcas de clase.

1.18 Con base en los datos de la tabla de frecuencias resultante del problema 1.13, determina las medidas de tendencia central.

1.19 Con base en los resultados de la tabla de frecuencias del problema anterior, construye el histograma y el polígono de frecuencias, ubicando y comprobando el valor de la moda.

1.20 Con base en los resultados del problema 1.15, calcula los valores de las medidas de tendencia central.

1.21 Con base en los resultados del problema 1.15, construye el histograma y el polígono de frecuencia.

1.22 Con base en los resultados del problema 1.15, construye las ojivas con el propósito de ubicar y comprobar el valor de la mediana.

1.23 Con base en los resultados del problema 1.17, determina las medidas de tendencia central.

1.24 Con base en los resultados de los problemas 1.17 y 1.23, construye el histograma y el polígono de frecuencias y ubica y comprueba el valor de la moda.

1.25 Con base en los resultados del problema 1.17, construye las ojivas, procediendo a ubicar y comprobar el valor de la mediana.

1.26 Con base en los resultados del problema 1.14, construye el histograma y el polígono de frecuencias, procediendo a ubicar y comprobar el valor de la moda.

1.27 Con base en los resultados del problema 1.15, elabora la tabla de frecuencias de manera que determine las medidas de tendencia central.

1.28 Con base en los resultados del problema 1.27, construye el histograma y el polígono de frecuencias, procediendo a ubicar y comprobar el valor de la moda.

1.29 Con base en los resultados del problema 1.28, desarrolla las ojivas de manera que ubiques y compruebes el valor de la mediana.

T **1.30** A partir de la colección de datos que se muestra, realiza un análisis de estadística descriptiva para datos no agrupados determinando el valor de las medidas de tendencia central.

169	128	140	176	161	119	164	154
140	136	154	136	148	144	175	148
136	162	130	141	123	159	151	150
146	139	160	146	136	148	144	152
145	142	156	145	128	150	135	161

1.31 Con base en la colección de datos del problema 1.30, determina el sexto decil y el 85avo percentil.

1.32 Con base en la colección de datos del problema 1.30, determina el número de intervalos aplicando la fórmula de Sturges, así como la amplitud de clase y la tabla de frecuencias.

1.33 Con base en los resultados del problema anterior, construye el histograma, el polígono de frecuencias y las ojivas.

1.34 Con base en la tabla de frecuencias del problema 1.30, calcula los valores del sexto decil y el 85avo percentil, determinando si existe diferencia o no con respecto a los resultados del problema 1.31.

1.35 Con base en la tabla de frecuencias del problema 1.30, determina la caracterización de la curva de distribución de frecuencias a través de un análisis descriptivo por momentos estadísticos.

1.36 Con base en la colección de datos que se muestra, calcula las medidas de tendencia central con base en datos no agrupados.

49	121	98	21	56	77
68	34	87	45	72	66
89	46	65	78	37	118
45	78	41	120	98	90
95	90	32	114	45	81

1.37 Con base en la colección del problema anterior, calcula Q_1, Q_2, Q_3, D_2, D_7, P_{48} y P_{85}.

1.38 Con base en la colección de datos que se muestra en el problema 1.36, determina el número de intervalos por regla empírica, procediendo a diseñar los mismos.

1.39 Con base en la colección de datos del problema 1.36, determina el número de intervalos por la fórmula de Sturges, procediendo a diseñar los mismos, y argumenta: ¿existe diferencia en el número de intervalos y diseño con los propuestos por el método empírico?

1.40 Con base en los resultados del problema 1.36, desarrolla la tabla de frecuencias de manera que determine las medidas de tendencia central.

1.41 Con base en los resultados del problema 1.40, construye el histograma y el polígono de frecuencias.

1.42 Con base en los resultados del problema 1.40, determina el valor de los siguientes cuantiles: D_1, D_4, D_7, Q_1, Q_3, P_{50}, P_{67} y P_{72}.

1.43 Con base en los resultados del problema 1.40, construye la ojiva ascendente y corrobora gráficamente los cuantiles: D_4, Q_3, P_{50} y P_{72}.

1.44 Con base en la tabla de frecuencias correspondiente a la colección de datos del problema 1.36, obtén la caracterización de la curva de distribución de frecuencias mediante un análisis por momentos estadísticos.

1.45 En relación con la tabla de frecuencias que se presenta en el problema 1.13, calcula el valor de la media geométrica.

Estadísticos e-Books & Papers

1.46 Con base en la tabla de frecuencias que se presenta en el problema 1.14, calcula el valor de la media geométrica.

1.47 Con base en la colección de datos del problema 1.17, compara el valor de la media aritmética con el valor de la media geométrica.

1.48 Considerando la colección de datos del problema 1.36, calcula la media geométrica por medio de la fórmula para datos no agrupados.

1.49 Compara el valor de la media aritmética con el valor de la media geométrica de datos agrupados del problema 1.36.

T **1.50** Usa la siguiente tabla de frecuencias a continuación para contestar lo que se te pide.

Tabla 1.29

Intervalos nominales		f
Límite inferior	Límite superior	
18	26	14
27	35	23
36	44	56
45	53	32
54	62	45

Continua tabla

Intervalos nominales		f
Límite inferior	Límite superior	
63	71	56
72	80	56
81	89	40
90	98	22
99	107	56

a) Determina el valor de las medidas de tendencia central.

b) Elabora el histograma y el polígono de frecuencias, así como la comprobación gráfica del valor de la mediana mediante el cruce de ojivas.

c) Determina las características de la curva de distribución de frecuencias mediante momentos estadísticos.

d) Determina el valor de los cuartiles y de todos los deciles.

e) Determina la proporción de datos que se ubican entre el 3er cuartil y el 95avo percentil; así como entre el 2° decil y la mediana.

f) Determina el valor de la media geométrica.

PROBLEMAS RETO

1

En una importante empresa de alimentos se realizan diferentes pruebas de laboratorio para cuidar la calidad de sus productos, una de ellas es verificar el peso de cada uno de sus productos. Cada lunes en el laboratorio se reciben cinco lotes y cada uno de ellos consta de 25 paquetes de 250 gr. Los registros que se obtienen se encuentran capturados en la siguiente tabla de frecuencias. El gerente del área de Control de Calidad le pide a su analista que le proporcione la siguiente información.

Tabla 1.30

Peso	Frecuencia
239	4
240	6
245	15
247	7
248	11
250	65
252	16
254	1

a) Determina el valor de las medidas de tendencia central.

b) Elabora el histograma y el polígono de frecuencias, así como la comprobación gráfica del valor de la mediana por medio del cruce de ojivas.

c) Determina las características de la curva de distribución de frecuencias mediante momentos estadísticos.

SR Problemas aplicados a la realidad T Problemas para resolver con tecnología

Estadísticos e-Books & Papers

d) Determina el valor de los cuartiles y de todos los deciles.

- Determina el valor de la media geométrica.
- ¿Crees que es necesario revisar las básculas?

2

Una empresa de cosméticos le encargó a una encuestadora determinar el grado de satisfacción de sus clientas por una crema desmaquillante que tiene en el mercado más de tres años. La respuesta se catalogó en: Excelente (E), Buena (B), Regular (R) o Mala (M). Los resultados obtenidos son los siguientes:

B	E	B	E	E	E	R	E	M
B	M	M	B	M	B	M	E	B
E	R	E	B	E	R	E	E	B
R	E	B	B	E	B	E	B	E
B	R	E	E	M	R	B	B	R
M	R	M	B	B	R	B	R	B
M	R	B	B	R	R	M	B	R
R	B	B	B	E	M	M	M	R
B	E	E	M	B	B	B	R	M
E	E	B	E	B	E	M	M	R
B	R	M	E	B	E	E	E	R
E	B	E	B	B	M	B	E	E
M	E	B	B	R	R	R	E	M
B	B	E	B	B	M	E	E	E
E	R	R	E	R	M	E	M	B
B	B	M	E	B	B	B	B	M

B	R	B	E	B	E	B	B	E
E	B	E	E	B	M	E	R	E
E	E	M	E	R	R	E	R	M
B	B	M	E	B	M	B	M	E
E	B	E	B	B	E	E	M	M
R	E	M	B	E	B	E	E	B
E	R	M	B	M	E	B	E	B
R	E	M	E	M	E	B	E	M
E	B	B	E	B	B	B	B	E
B	M	B	E	B	E	B	B	E
R	B	B	R	E	E	R	B	E
E	E	E	E	B	B	B	R	B
R	B	R	B	E	B	B	M	B
M	M	R	R	E	E	E	E	R
E	R	B	B	B	M	B	B	B
E	E	B	E	E	R	M	R	E

- Construye una tabla de frecuencias, si gustas puedes utilizar Excel.
- ¿Cuál es la ojiva?
- Elabora su histograma y su polígono de frecuencias.
- Determina la caracterización de la curva de distribución de frecuencias mediante un análisis descriptivo por momentos estadísticos.
- ¿Qué porcentaje de personas considera que la crema es mala?
- ¿Cuántas clientas consideran que la crema es excelente?
- ¿Cuál es la imagen general de la crema que tienen las personas?
- ¿Cuáles son tus conclusiones?

3

Petrolax es una importante empresa petrolera, que desea contratar el suministro de tubos de acero para sus instalaciones, solo que le piden que la asignación del proveedor la tiene que realizar por medio de una licitación abierta. Para la licitación se presentaron tres empresas (Tubular, Tubos de acero y Acero y tubos), las tres empresas venden la unidad al mismo precio y con las mismas especificaciones técnicas del material. Petrolax solicita que el proveedor mantenga un diámetro promedio por cada 20 tubos entregados de 46 pulgadas; para lo cual solicitó a cada empresa un lote muestra de este tamaño. El área de metrología de Petrolax obtuvo los siguientes resultados (las unidades están dadas en pulgadas).

487
Estadísticos e-Books & Papers

Tabla 1.31		
Tubular	Tubos de acero	Acero y tubos
46	45	47
46	47	46
45	47	46
46	46	46
45	46	45
46	47	46
44	47	46
46	45	45
45	46	46
44	46	46
46	46	46
45	46	45
45	45	46
46	46	47
44	46	46
46	46	45
45	46	46
45	45	46
44	45	46
45	45	45

¿Qué proveedor seleccionarías? Es muy importante justificar tu respuesta con un buen análisis.

REFERENCIAS

Berenson, Mark L. y David M. Levine (1996). *Estadística básica en administración* (6a. ed.). México: Prentice Hall Hispanoamericana.

Hines, William W., Douglas C. Montgomery, David M. Goldsman y Connie M. Borror. *Probabilidad y estadística para ingenieros* (4a. ed.). México: Patria, 3a. reimpresión.

Kohler, Heinz (1996). *Estadística para negocios y economía* (1a. ed.). México: Compañía Editorial Continental.

Lind, Douglas, William G. Marchal y Samuel A. Wathen (2005). *Estadística aplicada a los negocios y a la economía* (12a. ed.). México: McGraw-Hill.

Mendenhall, William (1999). *Estadística para administradores* (2a. ed.). México: Grupo Editorial Iberoamérica.

DIRECCIONES ELECTRÓNICAS

http://www.youtube.com/watch?v=pFotql8CRLU

http://www.slideshare.net/jeffertyni/histograma

Estadísticos e-Books & Papers

Estadísticos e-Books & Papers

UNIDAD 2

Teoría de la probabilidad y distribuciones de probabilidad

OBJETIVOS

- Aplicar el concepto de probabilidad y su trascendencia en la vida cotidiana y profesional.
- Entender los experimentos como actividades que dan origen a los eventos.
- Conocer y aplicar los diferentes tipos de probabilidad.
- Representar y operar los espacios muestrales a través de diagramas de Venn.
- Entender el concepto de variable aleatoria y su relación con la realidad.
- Distinguir las diferencias entre las variables aleatorias, discretas y continuas.
- Comprender y distinguir las principales diferencias entre distribución de probabilidad y función de densidad.
- Conocer y distinguir las diferencias entre las distribuciones de probabilidad discretas.
- Entender y distinguir el concepto y las características de una distribución continua.
- Entender y aplicar el teorema del límite central.

¿QUÉ SABES?

- ¿Cuál es la diferencia entre un experimento y un evento?
- ¿Qué es el espacio muestral?
- ¿Cómo apoya la teoría de conjuntos al análisis de la probabilidad?
- ¿Qué es una permutación y qué es una combinación?
- ¿Cuál es la diferencia entre una distribución de probabilidad y una función de densidad?
- ¿Cuáles son las principales diferencias entre las distribuciones binomial y de Poisson?
- ¿Cuáles son las principales características de la curva de distribución normal?
- ¿Qué son las unidades estandarizadas?

Estadísticos e-Books & Papers

2.1 Introducción

A lo largo de la historia, el hombre siempre ha deseado tener la certeza de los hechos que son de su interés. Por esa razón, ha recurrido a diversos medios, que involucran desde las predicciones, las cuales ofrecen criterios subjetivos, hasta la interpretación sobrenatural de los hechos. Sin embargo, el desarrollo de la teoría de la probabilidad ofrece el análisis de los hechos basados en la objetividad, los cuales se exponen a lo largo de este capítulo.

2.2 Concepto de probabilidad

Probabilidad se define como la cuantificación de la posibilidad de que un hecho ocurra.

Alerta

La probabilidad es la cuantificación de la posibilidad de que un hecho ocurra.

El análisis de la probabilidad se fundamenta en los experimentos, los cuales se definen como: *el conjunto de actividades interrelacionadas que permite el desarrollo de un fenómeno de manera parcial o total, y donde al resultado de un experimento se le denomina evento.*

Con base en la definición anterior, debe tenerse presente que un experimento puede tener más de un evento (resultado). Por tanto, al conjunto de eventos se le denomina *espacio muestral*, al cual, por lo general, se le denota con la letra *S* (del inglés *space*, que en español significa espacio).

A continuación se presenta una serie de ejemplos de experimentos y sus correspondientes espacios muestrales.

Alerta

Los eventos son los diferentes resultados de un experimento.

Ejemplos

- **Experimento:** Lanzar un dado $S = \{1, 2, 3, 4, 5, 6\}$
- **Experimento:** Lanzar una moneda $S = \{águila, sol\}$
- **Experimento:** Observación del género de los recién nacidos en un servicio de maternidad $S = \{varón, mujer\}$

2.3 Los espacios muestrales y la teoría de conjuntos

Alerta

Se denomina **espacio muestral** al conjunto de eventos.

Es importante resaltar que los espacios muestrales de un experimento se consideran conjuntos, donde los eventos son subconjuntos de los mismos.

Con base en lo expuesto antes, a los eventos de los espacios muestrales se les representa de acuerdo con la notación propia de los conjuntos, como se muestra en los ejemplos siguientes:

a) Denotación con una letra mayúscula, para identificar el conjunto, y listar entre llaves los eventos.
- **Experimento:** Lanzar un dado.
- **Espacio muestral:** $S = \{1, 2, 3, 4, 5, 6\}$

b) Denotación con una letra mayúscula, para identificar el conjunto y exponer entre llaves un argumento lógico.
- **Experimento:** Lanzar un dado.
- **Espacio muestral:** $S = \{X: 1 \leq X \leq 6\}$

c) Empleo de los llamados diagramas de Venn o Euler.

En este caso, el área delimitada por el rectángulo expresa al espacio muestral, mientras que el círculo adentro del rectángulo representa los eventos (véase figura 2.1).

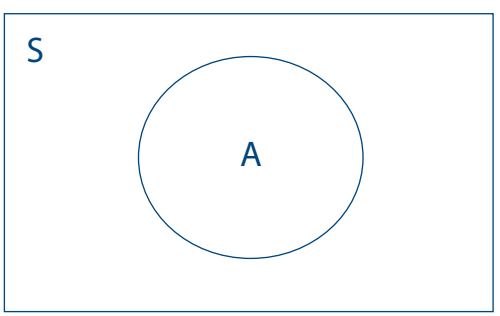


Figura 2.1

Estadísticos e-Books & Papers

Considerando el ejemplo en desarrollo:

- **Experimento:** Lanzar un dado.
- **Espacio muestral:** Está compuesto por seis eventos, por lo que el diagrama de Venn del espacio muestral queda expresado de la siguiente manera:

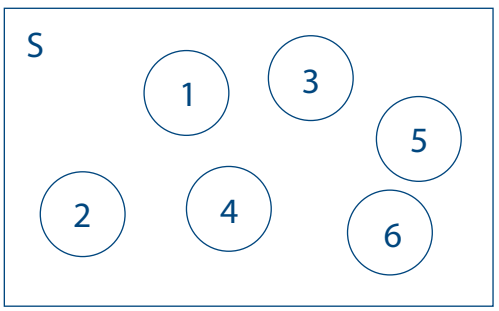


Figura 2.2

Operaciones con conjuntos

De manera clásica, algunas operaciones se identifican con los conjuntos. Estas operaciones y los conjuntos se aplican en el análisis de la probabilidad. A continuación se relacionan los más importantes:

a) **Unión de conjuntos.** Consiste en agrupar los elementos de los conjuntos en estudio en un solo bloque, en el cual simbólicamente se representan por $A \cup B$ (véase figura 2.3).

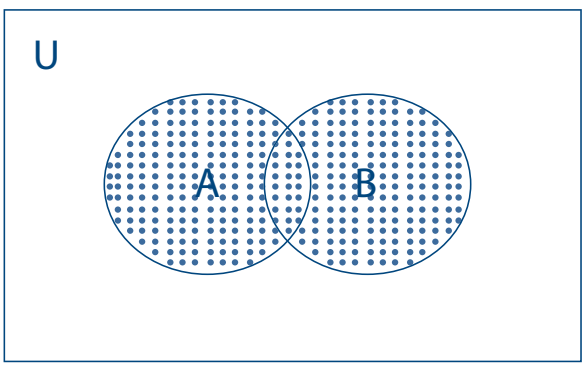


Figura 2.3

b) **Intersección de conjuntos.** Se refiere a considerar solo aquellos elementos que son compartidos por conjuntos en análisis; simbólicamente, la intersección de conjuntos se representa por $A \cap B$ (véase figura 2.4).

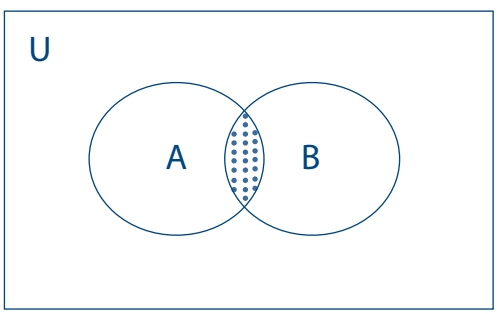


Figura 2.4

Estadísticos e-Books & Papers

Alerta

Los espacios muestrales se pueden representar mediante diagramas de Venn, donde los conjuntos simbolizan los eventos de un experimento.

c) **Conjunto complemento.** Este se compone por todos aquellos elementos que no forman parte de un conjunto. Simbólicamente, se representa como A^c o A' (véase figura 2.5).

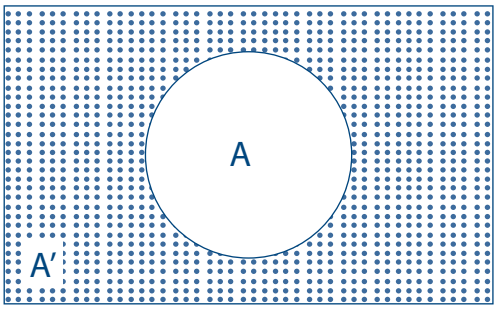


Figura 2.5

Problema resuelto

En favor de una campaña contra la obesidad, la cooperativa de una escuela secundaria ofrece jugos naturales de naranja y zanahoria. Al encuestar a 40 alumnos, 16 de ellos dijeron tomar jugo de zanahoria, 20 afirmaron tomar jugo de naranja, y 8 tomaron de los dos jugos.

En razón a las preferencias, determina:

- ¿Cuántos alumnos toman solo jugo de zanahoria?
- ¿Cuántos alumnos toman solo jugo de naranja?
- ¿Cuántos alumnos no toman ningún jugo?

Solución

Con base en el número de alumnos que dijeron tomar ambos jugos, primero se realiza la siguiente distribución:

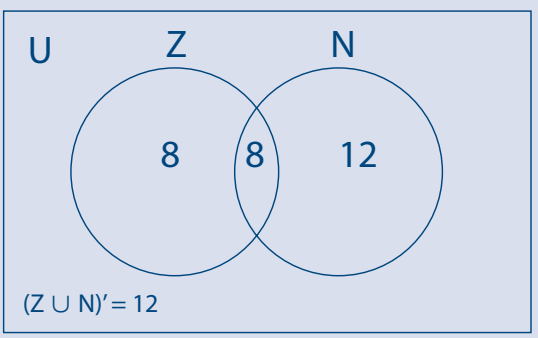


Figura 2.6

De la figura, obsérvese que los 8 alumnos que toman de los dos jugos forman parte de los conjuntos Z y N, de manera que el número de alumnos del conjunto Z es 16 (Z = 8 + 8). Lo mismo ocurre con el conjunto N, en donde el número de alumnos es 20 (N = 8 + 12). Sin embargo, al realizar el recuento de los alumnos, estos suman 28 entre los dos conjuntos, por lo que el complemento de la unión de ellos es la diferencia de los 40 alumnos, que es 12 [(Z ∪ N)' = 12.

2.4 Probabilidad clásica

La probabilidad clásica se representa mediante una razón, la cual guarda la siguiente estructura:

$$P(A) = \frac{Núm.\ éxitos}{Núm.\ total\ de\ eventos}$$

Estadísticos e-Books & Papers

Donde $P(A)$ se lee como:

La probabilidad (P) de que ocurra el evento 'A'.

Aquí, cabe resaltar que las probabilidades se pueden expresar en fracciones, en números decimales o en porcentajes. Pero, por lo común, el valor de la probabilidad se ubica entre 0 y 1; es decir, el valor máximo que puede tomar una probabilidad es 1.

Alerta

Las probabilidades se pueden expresar en decimales, porcentajes o fracciones dentro del rango entre 0 y 1.

Problema resuelto

Supón que se lanza un dado balanceado. Determina: ¿cuál es la probabilidad de que salga el 4?

Solución

Primero, se debe definir el evento:

a) 4

Luego, se establece el espacio muestral:

$$S = \{1, 2, 3, 4, 5, 6\}$$

Por tanto, como se puede observar, existe un solo evento que cumple con lo propuesto. Esto es, solo hay un 4 en seis posible eventos.

Entonces:

$$P(A) = \frac{1}{6} = 0.1666 = 16.66\%$$

2.5 Probabilidad de eventos mutuamente exclusivos o excluyentes

La probabilidad de eventos mutuamente exclusivos o excluyentes se presenta cuando la ocurrencia de un evento no afecta la ocurrencia de otro u otros eventos.

De acuerdo con la teoría de conjuntos, estos eventos se expresan de la siguiente manera:

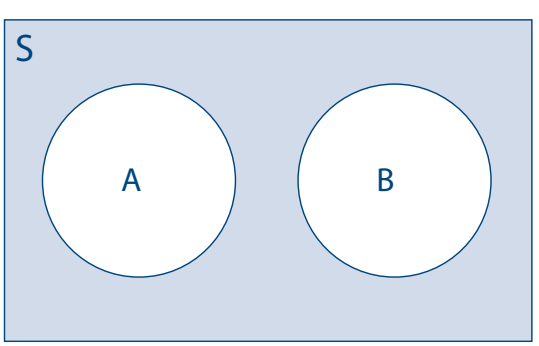


Figura 2.7

Donde el modelo matemático es el siguiente:

$$P(A \text{ o } B) = P(A) + P(B)$$

Para ejemplificar lo anterior, planteamos el siguiente problema.

Estadísticos e-Books & Papers

Problema resuelto

Supón que se lanza un dado. ¿Cuál es la probabilidad de que se obtenga 4 o 6?

Solución

De acuerdo con lo expuesto antes, primero se definen los eventos, que en este caso son dos:

A: Que se obtenga un 4.

B: Que se obtenga un 6.

Siendo el espacio muestral:

$$S = \{1, 2, 3, 4, 5, 6\}$$

Por tanto, las probabilidades de los eventos son:

$$P(A) = \frac{1}{6} \quad \text{y} \quad P(B) = \frac{1}{6}$$

En consecuencia:

$$P(A \text{ o } B) = P(A) + P(B) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

2.6 Probabilidad de eventos comunes

Alerta

La diferencia entre eventos mutuamente excluyentes y comunes, con respecto a su operación, solo se establece al definir los espacios muestrales correspondientes, con el fin de determinar la cantidad de eventos en común.

La probabilidad de eventos comunes, se refiere a la probabilidad donde parte de un evento *A* es parte de un evento *B*.

De acuerdo con la teoría de conjuntos, estos eventos quedan expresados de la siguiente manera:

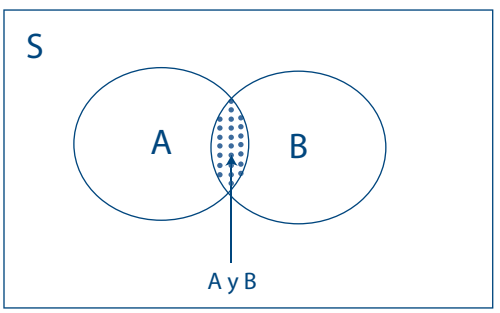


Figura 2.8

Donde el modelo matemático es el siguiente:

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$$

Nótese que el evento en común se encuentra tanto en *A* como en *B*, es decir está duplicado, por lo que debe restarse en una ocasión, de manera que solo sea contabilizado una sola vez.

Con el objeto de ejemplificar lo expuesto antes, se presenta el siguiente problema resuelto.

Problema resuelto

Como es sabido, una baraja cuenta con 52 cartas. Si en un juego de destreza se extrae una carta, determina: ¿cuál es la probabilidad de que esta sea una pica o un rey?

Estadísticos e-Books & Papers

Solución

Como en los casos anteriores, primero se definen los eventos:

A: Carta de pica.

B: Carta de rey.

Luego, se definen los espacios muestrales:

A = {2♠, 3♠, 4♠, 5♠, 6♠, 7♠, 8♠, 9♠, 10♠, J♠, Q♠, K♠, A♠}

B = {K♣, K♥, K♥, K♣}

Como se puede observar en los espacios muestrales, aquí también coexiste la carta del rey de picas (K♠), por lo que las probabilidades de cada evento son las siguientes:

$$P(A) = \frac{13}{52} \qquad P(B) = \frac{4}{52} \qquad P(A \text{ y } B) = \frac{1}{52}$$

Por tanto:

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

2.7 Probabilidad de eventos simultáneos o sucesivos

Este tipo de probabilidad se refiere al hecho de que, durante el desarrollo de un experimento, los eventos suceden al mismo tiempo o uno después del otro.

El modelo matemático que define la probabilidad de eventos simultáneos o sucesivos es el siguiente:

$$P(A \cap B) = P(A) \cdot P(B)$$

A continuación se expone un planteamiento clásico en el siguiente problema resuelto, para ejemplificar lo anterior.

Problema resuelto

En un experimento de azar, se lanza una moneda dos veces seguidas. Determina: ¿cuál es la probabilidad de que la moneda caiga dos veces consecutivas en águila?

Solución

En este caso, lo primero que debe hacerse es señalar que en este experimento, en específico, cada vez que se lanza la moneda, el espacio muestral es el mismo, independientemente del resultado del evento inmediato anterior. Esto se representa de la siguiente manera:

S = {Sol (S), Águila (A)}

Por tanto, en el primer lanzamiento, la probabilidad de que la moneda caiga sol es:

$$P(S) = \frac{1}{2}$$

Por su parte, en el segundo lanzamiento, la probabilidad de que la moneda caiga sol es:

$$P(S) = \frac{1}{2}$$

En consecuencia, el valor de la probabilidad es:

$$P(SS) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

Alerta

La probabilidad de eventos consecutivos o simultáneos se obtiene multiplicando las probabilidades de los eventos.

Estadísticos e-Books & Papers

2.8 Los diagramas de árbol

Alerta

Los diagramas de árbol exponen de manera gráfica la secuencia de eventos, así como las probabilidades de ocurrencia de los mismos.

Una de las herramientas que se utiliza más comúnmente para el análisis de la ocurrencia de los eventos y, en su caso, para estudiar aquellos problemas que requieren del análisis de una secuencia de eventos, son los *diagramas de árbol*, debido a que por medio de estos se pueden listar y organizar todas las opciones viables acerca de un problema o condición.

Los diagramas de árbol son un despliegue gráfico que expone la secuencia de los eventos, así como las probabilidades de ocurrencia de los mismos. En general, en el diagrama de árbol es posible representar los diversos eventos que pueden suceder, como ramas que parten de una bifurcación. Por tanto, el diagrama arbóreo reúne conjuntamente una secuencia de eventos y probabilidades. Para una mejor comprensión de este tema, en la siguiente figura se presenta el modelo general de un diagrama de árbol.

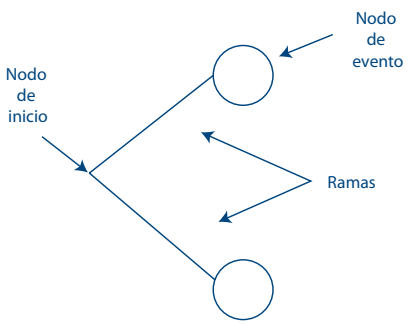


Figura 2.9

Para ejemplificar lo anterior considérese el siguiente problema resuelto.

Problema resuelto

En un experimento al azar, se lanza una moneda dos veces de manera continua. Construye el diagrama de árbol correspondiente, de tal manera que expongas el espacio muestral correspondiente.

Solución

Primero, es indispensable recordar que el espacio muestral del lanzamiento de una moneda es:

$$S = (Sol, Águila)$$

Por consiguiente, después de lanzar por primera vez la moneda, se tienen dos resultados posibles: sol y águila; no obstante, al lanzar la moneda de nuevo, independientemente del resultado de la primera experiencia, cada evento tiene dos eventos posibles: sol y águila (véase figura 2.10).

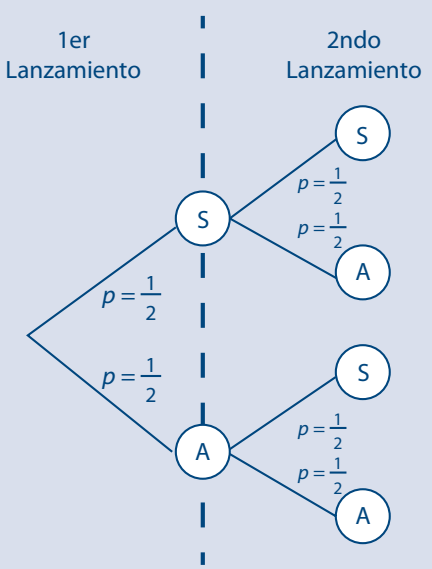


Figura 2.10

Estadísticos e-Books & Papers

2.9 Eventos consecutivos con y sin reemplazo

El desarrollo de algunos experimentos propone que estos se puedan repetir sucesivamente bajo ciertas condiciones que determinen que el espacio muestral cambie en cuanto al número de eventos.

De manera explícita, es posible considerar que los elementos que generan los eventos de un experimento puedan ser reconsiderados para la siguiente réplica del mismo o no. Esto se conoce como condiciones con reemplazo o sin reemplazo, como se observa en el siguiente ejemplo.

Ejemplo

Experimento: Seleccionar una carta de una baraja.

Con reemplazo: La carta se vuelve a integrar al mazo.

Sin reemplazo: Una vez seleccionada la carta es retirada y, por tanto, no es considerada en las réplicas posteriores.

Con base en las condiciones expuestas, es posible señalar que las probabilidades se conservan en el caso con reemplazo, lo cual no sucede en el caso sin reemplazo, donde estas cambian.

Para ejemplificar lo anterior, se retoma el caso clásico del bote con canicas.

Alerta

Los espacios muestrales de eventos consecutivos con reemplazo permiten conservar las probabilidades de ocurrencia a lo largo de las réplicas de las experiencias.

Problema resuelto

En un bote se colocan cinco canicas (tres blancas y dos negras); inmediatamente después, se agita el bote, con el fin de sortear su contenido. Determina: ¿cuál es la probabilidad de sacar de manera consecutiva una canica negra y una canica blanca con y sin reemplazo?

Solución

El problema plantea un experimento que se puede desarrollar bajo dos condiciones. De esta manera, en primera instancia se determina la probabilidad con reemplazo (véase figura 2.11).

En este caso, luego se definen los eventos:

A: Canica negra (*CN*)

B: Canica blanca (*CB*)

Considerando que en la primera extracción se obtiene la canica negra, la probabilidad de este evento es:

$$P(CN) = \frac{2}{5}$$

Para la réplica del experimento, se devuelve la canica negra al bote y se sacude para sortear nuevamente. Considérese que en esta segunda extracción se obtiene una canica blanca; en ese caso, la probabilidad asociada a este evento es:

$$P(CB) = \frac{3}{5}$$

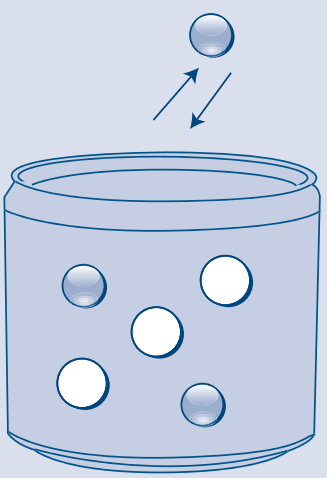


Figura 2.11

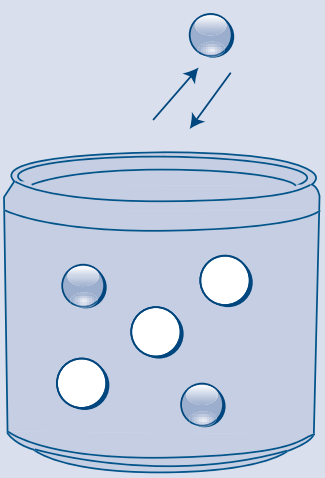


Figura 2.12

Estadísticos e-Books & Papers

Alerta

Los espacios muestrales de eventos consecutivos sin reemplazo son diferentes a lo largo de las réplicas de las experiencias, por lo que el valor de las probabilidades cambia.

Solución *(continuación)*

Por tanto, la probabilidad de estos eventos sucesivos es:

$$P(CN \cap CB) = \left(\frac{2}{5}\right)\left(\frac{3}{5}\right) = \frac{6}{25}$$

Para el caso sin reemplazo, se demuestra que las probabilidades cambian de un evento a otro, como se estudia más adelante.

2.10 Probabilidad condicional

Este tipo de probabilidad expone la ocurrencia de un evento *B*, una vez que ha ocurrido un evento *A*.

El modelo matemático que define lo anterior es el siguiente:

$$P(A \cap B) = P(A) \cdot P(B/A)$$

Donde *P(B/A)* se lee como: *"La probabilidad de B dada la ocurrencia de A"*.

Para ejemplificar lo anterior, retomamos el problema del bote con canicas en el siguiente problema resuelto.

Problema resuelto

Determina el valor de la probabilidad de extraer de manera consecutiva una canica blanca (*CB*) y posteriormente una canica negra (*CN*), en riguroso orden sin reemplazo.

Solución

En primer lugar se definen los eventos:

A: canica blanca *CB*

B: canica negra *CN*

De manera que: *B*/A

Como se puede observar, se señala la ocurrencia de *B* si y solo si ha ocurrido *A*; por consiguiente, las probabilidades de cada evento son:

$$P(A) = \frac{3}{5}$$

Considerando que se haya obtenido la canica blanca en la primera extracción, entonces el espacio muestral se ajusta y, en consecuencia, la probabilidad del evento condicional es:

$$P(B/A) = \frac{2}{4}$$

Por tanto:

$$P(A \cap B) = P(A) \cdot P(B/A) = \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20}$$

2.11 Teorema de Bayes

El teorema de Bayes fue desarrollado por el clérigo y matemático inglés Thomas Bayes, cuyos fundamentos fueron publicados *post mórtem*. De manera concreta, el teorema que lleva su nombre expone la reformulación de un conjunto de probabilidades previas (probabilidades *a priori*) como consecuencia

Estadísticos e-Books & Papers

de contar con información adicional acerca de los eventos del fenómeno estudiado, dando origen a nuevas probabilidades denominadas probabilidades *a posteriori*.

El modelo matemático del teorema de Bayes se fundamenta en la probabilidad condicional, como se muestra a continuación:

$$P(A/B) = \frac{P(A) \cdot P(B/A)}{P(B)}$$

De acuerdo con esta expresión matemática, se puede interpretar que el evento condicional (*B/A*) es parte de un evento *A*; por tanto, se procede a establecer la proporción de la probabilidad que ocupa el evento *B/A* al interior del evento *A*.

Para ejemplificar lo anterior considérese el siguiente problema resuelto.

Problema resuelto

Una planta productora de gelatinas cuenta con tres máquinas empacadoras. Así, la distribución del volumen de empaque se realiza de la siguiente manera:

- Máquina 1: 38%
- Máquina 2: 32%
- Máquina 3: 30%

De esta manera, la probabilidad de que el empaque salga defectuoso es de 11%, 15% y 14%, respectivamente por cada máquina.

La gerencia de producción de la planta está interesada en conocer cuál es la probabilidad de que si se selecciona una unidad al azar y es defectuosa, esta se haya empacado en la máquina 2.

Solución

Para la resolución de este problema, primero se precisa definir los eventos. Así, de estos, el primero lo constituye la máquina y el segundo es que el producto sea defectuoso. Por tanto, el evento condicional es que sea defectuoso dado que haya sido producido por la máquina 2.

Para ejemplificar de manera gráfica lo anterior, se sugiere desarrollar el diagrama de árbol correspondiente.

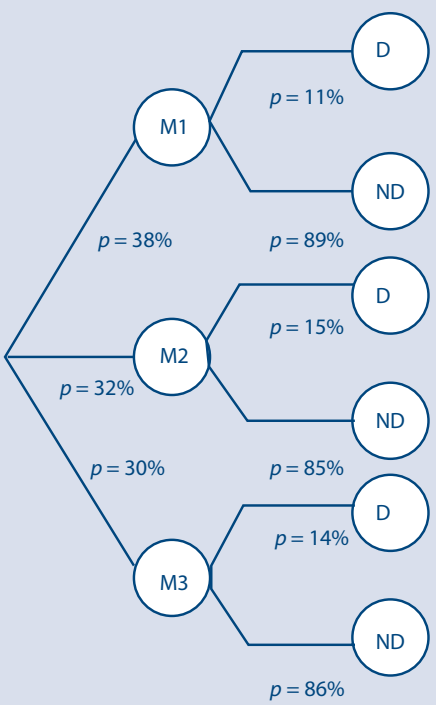


Figura 2.13

Estadísticos e-Books & Papers

Solución *(continuación)*

Luego, debe considerarse que las unidades defectuosas pueden provenir de *M1*, *M2* o *M3*, por lo que a estas se les considera eventos mutuamente exclusivos; esto es que el que una unidad haya sido producida en una máquina en específico no afecta el hecho de que provenga de otras máquinas. Entonces:

PM: Máquina

D: Unidad defectuosa

$$\therefore P(D) = P(PM1) \cdot P(D/PM1) + P(PM2) \cdot P(D/PM2) + P(PM3) \cdot P(D/PM3)$$

Tabla 2.1

Máquina	Producción (*PM*)	Defectuosa (*D*)	(*PM*) (*D*)
1	38%	11%	0.0418
2	32%	15%	0.0480
3	30%	14%	0.0420
		P(D)= $\sum$	0.1318

$$\therefore P(M2/D) = \frac{P(PM2) \cdot P(D/M2)}{P(D)} = \frac{(0.32)(0.15)}{0.1318} = 0.3642 = 36.42\%$$

2.12 El principio de multiplicación

El principio de multiplicación sucede cuando, en primera instancia, "algo" se puede llevar a cabo en "a" formas, en una segunda ocasión en "b" formas y en una tercera en "c" formas, y así sucesivamente. Por tanto, las "*n*" formas hechas en conjunto pueden ser expresadas como:

$$a \times b \times c \times \ldots \ (n \text{ factores}) \text{ formas}$$

Lo anterior queda ejemplificado en el siguiente problema resuelto.

Problema resuelto

Una fábrica de juguetes, la cual es la responsable de producir la muñeca de moda, ha diseñado un kit de guardarropa para esta muñeca, el cual está compuesto de tres vestidos: un azul, un gris y uno negro; así como también de dos pares de zapatos: un par de color rojo y un par de color amarillo.

¿Cuántas combinaciones de ropa para esta muñeca se pueden lograr con este kit de guardarropa?

Solución

Como primer paso, se sugiere desarrollar el diagrama de árbol, con el fin de mostrar todas las opciones del guardarropa desarrollado.

Como se puede observar del diagrama de árbol, existen seis formas posibles de combinar el guardarropa de la muñeca. Así, aplicando el principio de multiplicación, se sabe que existen tres formas de seleccionar el vestido (a) y 2 formas de seleccionar los zapatos (b). Por tanto:

$$a \times b = 3 \times 2 = 6 \text{ formas}$$

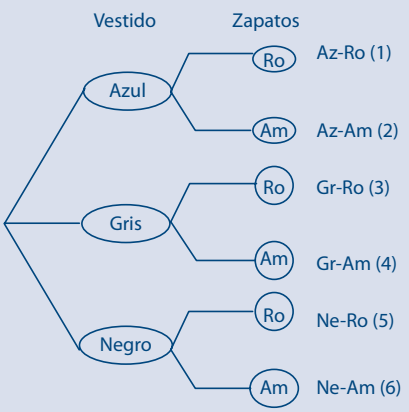


Figura 2.14

Estadísticos e-Books & Papers

2.13 Permutaciones

Una permutación es un arreglo donde los elementos que lo integran y su orden no importan.

Considérese el conjunto {a, b, c, d}. ¿Cuántas permutaciones de tres elementos pueden obtenerse de este conjunto?

La respuesta a esta pregunta puede responderse desarrollando el diagrama de árbol correspondiente:

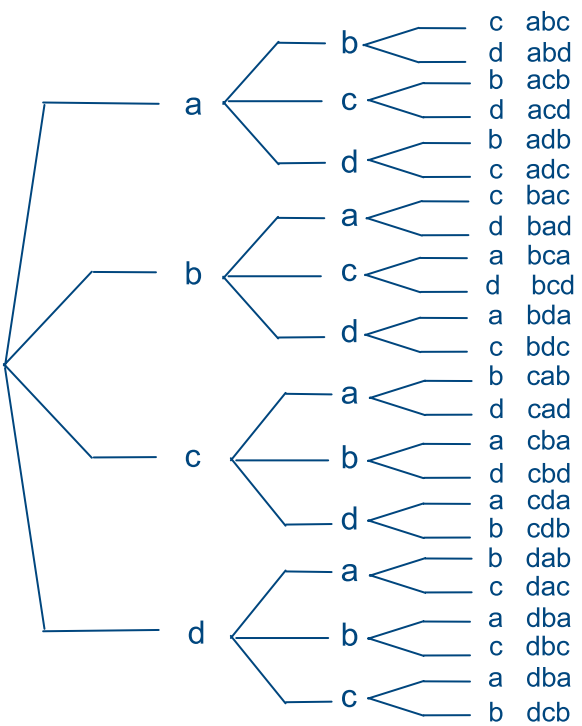


Figura 2.15

Como se puede observar del diagrama de árbol, si $k = 4$, $l = 3$ y $m = 2$; entonces:

$$k \cdot l \cdot m = (4)(3)(2) = 24 \text{ arreglos}$$

Nótese que el valor de $k = (n - 1)$ formas, $l = (n - 2)$ formas y $m = (n - 3)$ formas.

Por tanto, de manera general, se puede decir que el número de permutaciones de "n" elementos, tomados a la vez como permutaciones de un conjunto con "n" elementos, queda matemáticamente definido como:

$$_nP_n = n!$$

Aunque, si se considera que se pueden formar arreglos de "r" elementos, tomados a la vez a partir de un conjunto de "n" elementos, se tiene que la serie referida al cálculo queda expresada como:

$$(n)\,(n-1)\,(n-2)\,(n-3)\ldots(n-(r-1))$$

Así, la fórmula para determinar el número de permutaciones se expresa de la siguiente manera:

$$_nP_r = \frac{n!}{(n-r)!}$$

Para ejemplificar lo antes expuesto, a continuación se presenta el siguiente problema resuelto.

Alerta

Una permutación es un arreglo donde los elementos y su orden no importan.

Alerta

El cálculo del número de permutaciones se fundamenta en el principio de la multiplicación.

Estadísticos e-Books & Papers

Problema resuelto

Considerando el planteamiento del conjunto {a, b, c, d}, comprueba que el número de permutaciones de tres elementos tomados a la vez es 24.

Solución

Primero, se aplica la fórmula para el cálculo del número de permutaciones.

Así, si $n = 4$ y $r = 3$, entonces:

$$ {}_4P_3 = \frac{4!}{(4-3)!} = \frac{4 \times 3 \times 2 \times 1}{1} = 24 \text{ arreglos} $$

2.14 Combinaciones

Una combinación es un arreglo donde los elementos que lo integran y su orden sí importan.

Considérese de nueva cuenta al conjunto {a, b, c, d}. ¿Cuántas combinaciones de tres elementos pueden obtenerse del conjunto?

Con base en el diagrama de árbol desarrollado antes, se tiene:

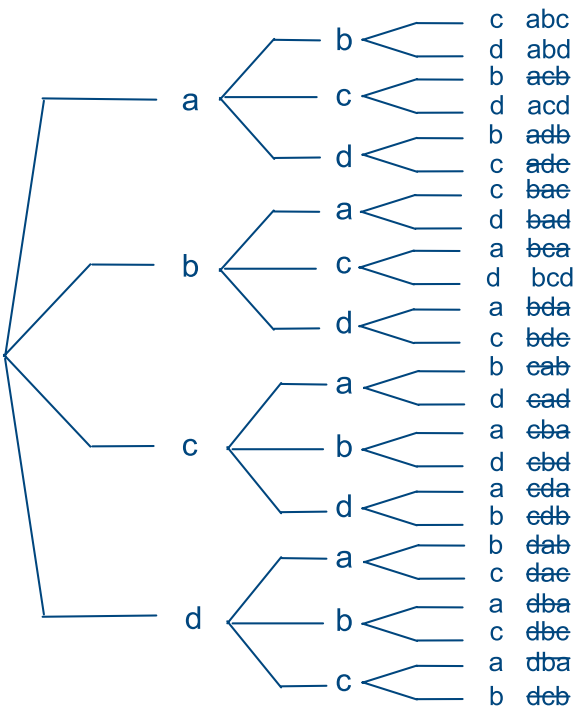


Figura 2.16

Alerta

Una combinación es un arreglo donde los elementos y su orden **sí** importan.

Obsérvese que el arreglo "abc" es igual a los arreglos "bac", "bca", "cab" y "acb", por lo que únicamente se considera el primer arreglo. Entonces, siguiendo con el criterio anterior, las combinaciones a considerar son cuatro: "abc", "abd", "acd" y "bcd".

Para calcular el número de combinaciones, la fórmula de las permutaciones debe ajustarse considerando el número de arreglos de "r" elementos:

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{r!\,(n-r)!} $$

Alerta

Una combinación siempre es diferente a una permutación.

Estadísticos e-Books & Papers

Para demostrar lo anterior, se expone el siguiente problema resuelto.

Problema resuelto

Considerando nuevamente el planteamiento del conjunto {a, b, c, d}, comprueba que el número de combinaciones de tres elementos, tomados a la vez, es cuatro.

Solución

Aplicando la fórmula para el cálculo del número de combinaciones, si $n = 4$ y $r = 3$, entonces:

$$ {}_4C_3 = \frac{4!}{3! \cdot (4-3)!} = \frac{4 \times 3 \times 2 \times 1}{3 \times 2 \times 1} = 4 \text{ arreglos} $$

2.15 Variables aleatorias

En general, una variable es un hecho cuantificable relacionado con un fenómeno; en particular, una variable aleatoria constituye un hecho cuantificable asociado a un experimento, cuyos valores están determinados por el azar dentro de los números reales. En otras palabras, se puede señalar que existe una relación funcional entre los eventos que conforman el espacio muestral del experimento y los números reales.

Alerta

Las variables aleatorias son hechos medibles relacionados con un experimento y cuyo valor está determinado por el azar.

De esta manera, las variables aleatorias se clasifican en:

- **Variables aleatorias discretas.** Si los valores asociados a los eventos de un experimento se obtienen mediante conteo, dichos valores se determinan de manera puntual dentro del terreno de los enteros. Es decir, sus posibles valores son numerables dentro de espacios finitos o infinitos.
- **Variables aleatorias continuas.** Si los valores asociados a los eventos de un experimento se obtienen, principalmente, de procesos de medición, sus valores se determinan dentro de un intervalo perteneciente a los reales.

2.16 Distribuciones de probabilidad

Una distribución de probabilidad es una función matemática real, la cual permite calcular el valor de la probabilidad de que ocurra un evento propio del espacio muestral de un experimento; en otras palabras, una distribución de probabilidad permite establecer las probabilidades de que se den los diferentes valores de una variable aleatoria.

Así, de manera concreta, las distribuciones de probabilidad se definen como un espacio de probabilidad y un espacio medible, donde una aplicación entre el espacio de probabilidad y el espacio medible es una variable aleatoria, siempre y cuando esta sea una aplicación medible.

Alerta

Las variables aleatorias se clasifican en discretas y continuas.

Las distribuciones de probabilidad se dividen en discretas y continuas, las cuales se analizan con detalle a continuación.

■ Distribuciones de probabilidad discretas

En este caso, es importante tener presente que las **distribuciones de probabilidad discretas** están relacionadas con fenómenos cuyos eventos (resultados) pueden tener origen en procesos de conteo, como los que se citan a continuación:

- Número de piezas defectuosas.
- Número de retardos.
- Número de empleados con incapacidad médica.

Estadísticos e-Books & Papers

Las distribuciones de probabilidad discreta se distinguen, entre otras cosas, por:

a) Contar con una función de densidad de probabilidades, la cual permite determinar el valor puntual de la probabilidad para un valor X de la variable aleatoria.

b) Contar con una función de distribución, la cual permite calcular la probabilidad acumulada; esto es, que permita obtener el valor de la probabilidad de un rango de valores de la variable aleatoria.

Alerta

Las variables aleatorias discretas cuentan con una función de densidad, la cual permite calcular un valor puntual de probabilidad.

Distribución binomial o de Bernoulli

Este tipo de distribución se diferencia por tener un espacio muestral basado en el éxito o el fracaso de generar un evento en un determinado número de ensayos.

Este tipo de distribución se fundamenta en el proceso de Bernoulli, el cual expone los siguientes puntos:

a) Cada prueba tiene un número concreto de resultados.

b) La probabilidad de éxito permanece constante en cada prueba.

c) Las pruebas son independientes, por lo que se pueden acumular sus resultados.

El modelo matemático que identifica este tipo de distribución es el siguiente:

$$P(r,\ n,\ p) = \frac{n!}{r! \cdot (n-r)!}\ p^r \cdot q^{n-r}$$

donde

Alerta

Las variables aleatorias discretas cuentan con una función de distribución, la cual permite calcular la probabilidad acumulada.

r = Número de éxitos.

n = Número de ensayos.

p = Probabilidad de éxito.

q = Probabilidad de fracaso: $q = 1 - p$.

Para ejemplificar lo anterior, considérese el siguiente problema resuelto.

Problema resuelto

En cierta región de nuestro país, se observa que las mujeres en edad fértil tienen, en promedio, tres hijos. De estos, la probabilidad de que alguno de ellos sea mujer es de 60%. Un estudio social desea determinar la probabilidad de que alguno de los tres hijos de cada una de estas mujeres sea varón.

Alerta

Las variables aleatorias discretas de aplicación común son del tipo binomial y de Poisson.

Solución

En este caso, es posible resolver el problema por medio de dos métodos de cálculo. En primera instancia, se puede solucionar a través del árbol de decisión, en el cual se aprecian de manera explícita los eventos que cumplen con la condición de que uno de los tres hijos sea varón; es decir, eventos consecutivos e independientes. El segundo método que se utiliza para resolver el problema es por medio de la fórmula de la distribución binomial.

El árbol de decisión de este problema se observa como el que aparece a continuación:

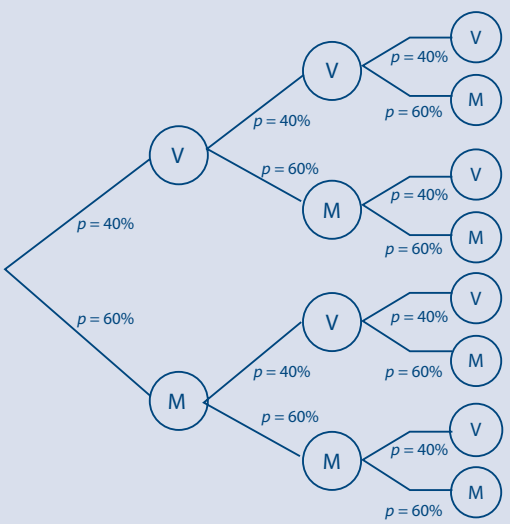


Figura 2.17

Estadísticos e-Books & Papers

Solución *(continuación)*

Como se puede observar de la figura, VMM, MVM o MMV son eventos independientes uno del otro; por tanto, el valor de la probabilidad es:

$$P(\text{VMM o MVM o MMV}) = P(\text{VMM}) + P(\text{MVM}) + P(\text{MMV})$$
$$= (0.4 \times 0.6 \times 0.6) + (0.6 \times 0.4 \times 0.6) + (0.6 \times 0.6 \times 0.4)$$
$$= 0.432 = 43.2\%$$

Por otra parte, aplicando la fórmula de la distribución binomial, donde:

$r = 1$
$n = 3$
$p = 0.4$
$q = 1 - 0.4 = 0.6$

Por tanto:

$$P(1,\ 3,\ 0.4) = \frac{3!}{1!(3-1)!}(0.4)^{1} \cdot (0.6)^{3-1} = 0.432 = 43.2\%$$

Ahora, para entender el concepto de función de distribución considérese el siguiente problema resuelto.

Problema resuelto

De acuerdo con el problema anterior (acerca de las mujeres en edad fértil), los responsables del estudio social requieren conocer los siguientes aspectos:

a) ¿Cuál es el valor de la probabilidad de que una mujer tenga cuando menos 2 varones?
b) ¿Cuál es la probabilidad de que una mujer tenga a lo más 2 varones?
c) ¿Cuál es la probabilidad de que una mujer tenga entre 2 y 3 varones?

Solución

Para la resolución de este problema, primero se calculan los valores de la probabilidad de que nazcan 0, 1, 2 y 3 varones y se transcriben en una tabla:

R	*P(r, n, p)*
0	21.60%
1	43.20%
2	28.80%
3	6.40%

Acto seguido, se procede a solucionar cada uno de los incisos planteados.

a) El término cuando menos 2, debe interpretarse como $X \geq 2$; esto es:

$P(X \geq 2) = 1 - [P(X=0) + P(X=1)] = 1 - [21.60\% + 43.20\%] = 35.20\%$

b) El término a lo más 2, debe interpretar como $X \leq 2$; esto es:

$P(X \leq 2) = P(X=0) + P(X=1) = 21.60\% + 43.20\% = 64.80\%$

c) El término entre dos y tres varones significa que $2 \leq X \leq 3$; por tanto:

$P(2 \leq X \leq 3) = P(X \leq 3) - P(X \leq 2) =$

$= [P(X=0) + P(X=1) + P(X=2)] - [P(X=0) + P(X=1)] =$

$= 93.60\% - 64.80\% = 28.80\%$

Estadísticos e-Books & Papers

Distribución de Poisson

Alerta

En la distribución de Poisson, los éxitos por unidad de observación son aleatorios; los eventos independientes se distribuyen de manera uniforme en la unidad de observación.

Este tipo de distribución de probabilidad discreta se distingue por considerar el número promedio de éxitos por unidad de tiempo, volumen o área. Su fórmula general es:

$$P(x) = \frac{\mu^x \cdot e^{-\mu}}{X!} = \frac{\lambda^x \cdot e^{-\lambda}}{X!}$$

donde

$\mu = \lambda$: Número promedio de éxitos.

X: Número de éxitos requerido.

Es importante destacar que la distribución de Poisson se fundamenta en las siguientes premisas:

a) La probabilidad de que ocurra un evento por unidad de tiempo, área o volumen es pequeña.

b) La probabilidad de que ocurran dos o más eventos de manera simultánea es prácticamente cero.

c) El número de eventos que ocurren por unidad de tiempo, área o volumen es independiente de otras observaciones, ya que no se encuentran relacionadas; es decir, son independientes.

De hecho, de manera similar a la distribución binomial, en la distribución de Poisson se pueden determinar las probabilidades de manera puntual o acumulada.

Para ejemplificar y exponer su aplicación se presenta el siguiente problema resuelto.

Problema resuelto

Una fábrica dedicada a la fabricación de duela de madera de cedro encuentra, durante un proceso de control de calidad, que en promedio hay dos defectos en la duela por cada metro cuadrado. La gerencia de control de calidad está interesada en saber las probabilidades de obtener:

a) Cuatro defectos exactamente.

b) Más de tres defectos.

c) A lo más tres defectos.

Solución

En primera instancia, se calculan las probabilidades de uno a cuatro defectos, aplicando la fórmula de Poisson:

X	*P(X)*
0	13.53%
1	27.07%
2	27.07%
3	18.04%
4	9.02%

De acuerdo con los resultados de la tabla, se procede a realizar los cálculos de cada inciso.

a) Para $X = 4$ la probabilidad es:

$$P(X=4) = 9.02\%$$

b) Para el caso de más de tres defectos, debe interpretarse una probabilidad cuando $P(X > 3)$:

$$P(X > 3) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)] = 1 - 0.8571 = 14.29\%$$

c) Para el caso de a lo más tres defectos, se debe entender que $P(X \leq 3)$. Por tanto:

$$P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = 85.71\%$$

Estadísticos e-Books & Papers

Aproximación a la distribución binomial a través de la distribución de Poisson

Toda distribución de probabilidad discreta tiene una media y una desviación estándar, las cuales se definen como:

$$\mu = np$$

y

$$\sigma = \sqrt{npq}$$

Con base en lo anterior, es posible señalar que en algunos problemas propuestos de distribución binomial se puede calcular la probabilidad ajustando la función de Poisson; aun cuando la probabilidad de éxito solo sea muy pequeña y se cuente con un número de ensayos grande.

$$P(r,\ n,\ p) = \frac{n!}{r!(n-r)!}\ p^r \cdot q^{n-r} = \frac{(n \cdot p)^X \cdot e^{-(n \cdot p)}}{X!}$$

Alerta

Las distribuciones de probabilidad discreta tienen una media y una desviación estándar definidas por: $\mu = np$ y $\sigma = \sqrt{npq}$.

Problema resuelto

De acuerdo con un estudio realizado con antelación, una compañía dedicada a la fabricación de equipos de cómputo encontró que de cada 20 000 discos duros, solo 19 880 discos cumplen con las especificaciones de control de calidad establecidas por la compañía. Si los lotes de embarque están compuestos por 600 unidades, la gerencia de control de calidad precisa saber si los valores de las probabilidades puntuales de hallar 0, 1, 2, 3, 4, 5 y 6 discos defectuosos en un lote son iguales entre la binomial y la aproximación de la binomial por Poisson.

Solución

$$\textit{Discos no defectuosos} = \frac{19\,980}{20\,000} = 99.40\%$$

Como en otros casos, primero se determina la probabilidad de que un disco sea defectuoso:

$$\therefore P(\textit{Disco defectuoso}) = 1 - P(\textit{Disco no defectuoso}) = 1 - 0.9940 = 0.006 = 0.6\%$$

donde

$P = 0.006$
$q = 0.994$
$n = 600$

Luego, se aplica la binomial:

R	*P(r,n,p)*
0	2.70%
1	9.79%
2	17.70%
3	21.29%
4	19.18%
5	13.80%
6	8.26%

Sin embargo, aplicando la de Poisson, se tiene:

$\mu = np = 600 \times 0.006 = 3.6$

Estadísticos e-Books & Papers

Solución *(continuación)*

X	P(X)
0	2.73%
1	9.84%
2	17.71%
3	21.25%
4	19.12%
5	13.77%
6	8.26%

Como se puede observar, las diferencias son mínimas entre los valores.

2.17 Curva de distribución de probabilidad discreta

Alerta

Las curvas de distribución de probabilidades muestran el comportamiento, en cuanto al cambio en los valores de las probabilidades, en razón del número de éxitos.

Los valores de la probabilidad de los eventos en análisis, a través de una distribución de probabilidad discreta, pueden exponerse de manera gráfica. De esta manera, al unir los valores que representan las probabilidades, se obtiene la curva de probabilidad correspondiente. Esta curva permite mostrar el comportamiento que sucede respecto del cambio en los valores de las probabilidades, en razón del número de éxitos.

Así, una distribución de probabilidad indica toda la gama de valores que pueden representarse como resultado de un experimento, si este se llevara a cabo.

Para mayor comprensión del tema, a continuación se expone el siguiente problema resuelto:

Problema resuelto

Con base en los datos del problema expuesto acerca de la probabilidad de nacimientos de varones en cierta región del país, desarrolla la curva de probabilidad correspondiente, sabiendo que el número mínimo y máximo de hijos varones por madre es de 0 y 3, respectivamente.

Solución

Primero, se calculan los valores de la probabilidad de que nazcan niños varones, que es de: 0, 1, 2 y 3. Enseguida, se procede a graficar estos datos, con el propósito de definir su curva de distribución.

r	P(r, n, p)
0	21.60%
1	43.20%
2	28.80%
3	6.40%

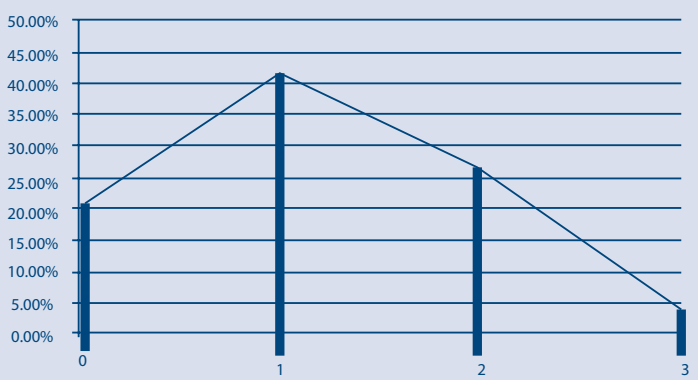


Figura 2.18

Estadísticos e-Books & Papers

2.18 Valor esperado de una variable aleatoria discreta o esperanza matemática

El valor que puede tomar una variable aleatoria discreta depende de su probabilidad de ocurrencia; por ello, algunos valores tienen menos probabilidades de ocurrir que otros. De acuerdo con lo expuesto antes, puede surgir la inquietud de saber: ¿cuál es el valor promedio de una distribución de probabilidades? Lo anterior da origen al concepto del valor esperado de una variable aleatoria discreta o esperanza matemática o promedio ponderado, la cual se estructura como el acumulado del producto de los valores que puede tomar la variable aleatoria y su probabilidad de ocurrencia correspondiente:

$$E(X) = \sum_{i=1}^{n} P_i \cdot X_i$$

donde

$E(X)$: Valor esperado o esperanza matemática ($E(X) \approx \mu$).

P_i: Probabilidad de ocurrencia del evento X_i.

X_i: Valor del evento "i".

Es preciso destacar aquí que la esperanza matemática se aproxima al promedio aritmético. De esta manera, al considerar un promedio es conveniente considerar la medición del grado de dispersión con respecto a la esperanza matemática, donde también resulta importante el hecho de considerar la probabilidad de ocurrencia de cada evento, por lo que la fórmula de la desviación estándar se modifica, como se muestra a continuación:

$$\sigma_p = \sqrt{\sum_{:=1}^{n} P_i \cdot [X_i - E(X)]^2}$$

En algunos cálculos, la probabilidad de ocurrencia se sustituye por el grado de importancia que un valor de la variable aleatoria tiene dentro de la distribución de probabilidad; como es el caso del costo porcentual promedio de capital o cuando en algún sorteo un cupón se repite varias veces en la urna.

Para la comprensión de lo anterior, obsérvese con cuidado el siguiente problema resuelto.

Problema resuelto

El encargado de una tienda que vende al mayoreo estima la probabilidad de ocurrencia de los diferentes niveles de venta de cajas de leche en polvo durante los cinco días de la semana. ¿Cuál es el valor esperado del nivel de ventas semanal? ¿Cuál el valor de la desviación estándar ponderada?

Tabla 2.2

Día	Ventas (Cajas)	Probabilidad
Lunes	180	20%
Martes	220	25%
Miércoles	260	20%
Jueves	240	15%
Viernes	200	20%

Solución

De acuerdo con la fórmula de la esperanza matemática, en primera instancia se procede como sigue:

Tabla 2.3

Día	Ventas (X)	Probabilidad (P)	
Lunes	180	20%	36
Martes	220	25%	55
Miércoles	260	20%	52
Jueves	240	15%	36
Viernes	200	20%	40
	$\sum =$	100%	219

Estadísticos e-Books & Papers

Solución *(continuación)*

$\therefore E(X) = 219$ cajas

Donde el valor de la media aritmética es de 219 cajas, con lo cual se confirma que la esperanza matemática se aproxima a la media aritmética.

De esta manera, el valor de la desviación estándar ponderada es:

Tabla 2.4

Día	Ventas (X)	Probabilidad (P)		$P[X - E(x)]^2$
Lunes	180	20%	36	304.20
Martes	220	25%	55	0.25
Miércoles	260	20%	52	336.20
Jueves	240	15%	36	66.15
Viernes	200	20%	40	72.20
	$\Sigma =$	100%	219	779.00

$$\sigma_p = \sqrt{779} = 27.91$$

2.19 Distribuciones de probabilidad continuas

Como se refirió antes, las distribuciones de probabilidad continuas son aquellas cuyos eventos se generan por procesos de medición y cuyos valores pertenecen a los números reales.

Las distribuciones de probabilidad continuas se relacionan con la denominada curva de *distribución normal* o *campana de Gauss*, la cual constituye una curva, cuyas principales características son las siguientes:

a) Su media vale 0 y el valor de su desviación estándar es 1.

b) Es simétrica alrededor de la media, por lo que la mediana y la moda tienen el mismo valor y ubicación que la media.

c) Su rango en unidades estandarizadas (Z) se encuentra entre -3 y $+3$; de hecho, los puntos de inflexión de la curva se dan para $\mu \pm \sigma$.

d) Su amplitud cubre $-\infty \leq x \geq +\infty$, de manera asintótica al eje x.

e) El valor del área bajo la curva es 1, ya que cubre todos los posibles eventos relacionados con una variable aleatoria.

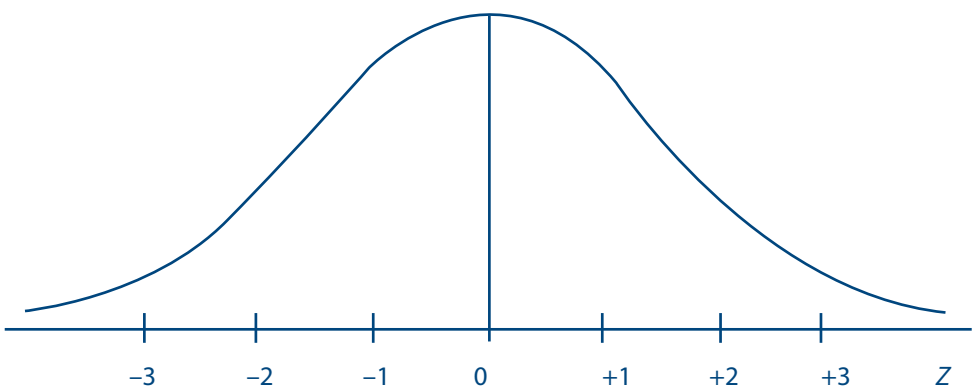


Figura 2.19

Estadísticos e-Books & Papers

La curva de distribución normal tiene la siguiente función de densidad:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Por esta razón, el cálculo de las probabilidades corresponde a la determinación del área bajo la curva delimitada por dos valores. Donde, para mayor referencia, se considera a partir de la media ($\mu = 0$) y un valor x, haciéndose necesario integrar la función de densidad de la distribución:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx$$

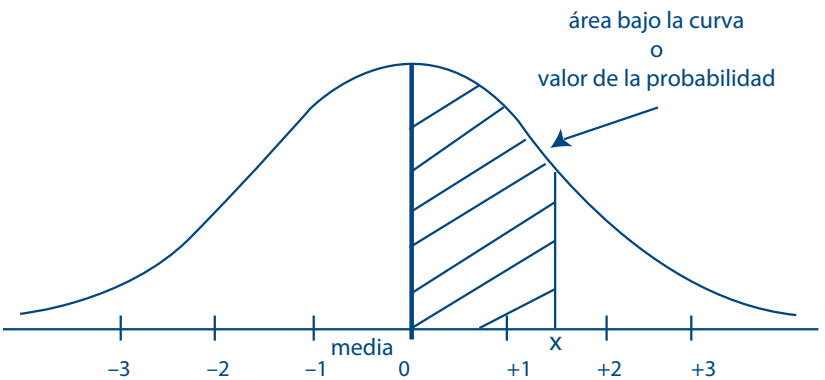


Figura 2.20

No obstante, resulta necesario hacer hincapié en que las áreas bajo la curva normal ya se encuentran tabuladas y que tan solo se requiere convertir las unidades de la variable aleatoria a unidades estandarizadas (Z), mediante de la siguiente fórmula:

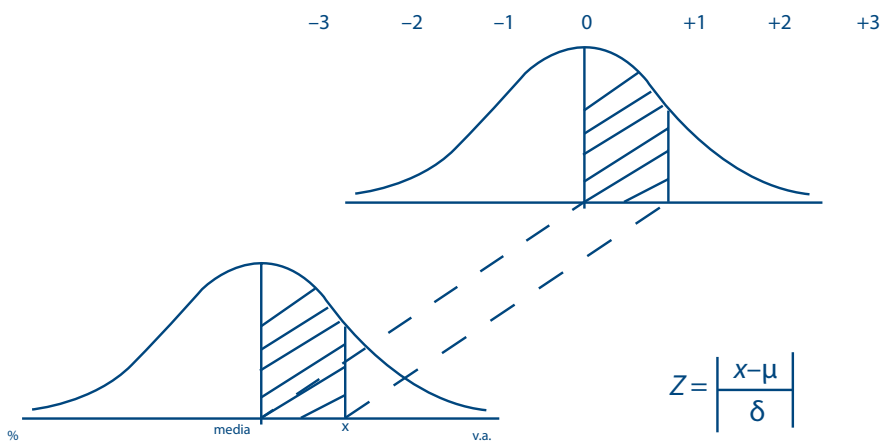


Figura 2.21

Si se desarrolla el cálculo, resulta posible apreciar las proporciones de la distribución que cubre, de acuerdo con la siguiente tabla y gráfica:

$\mu \pm \sigma$	68.20%
$\mu \pm 2\sigma$	95.40%
$\mu \pm 3\sigma$	99.70%

Alerta

La fórmula de Z expresa la proporción de σ que es $X-\mu$.

Estadísticos e-Books & Papers

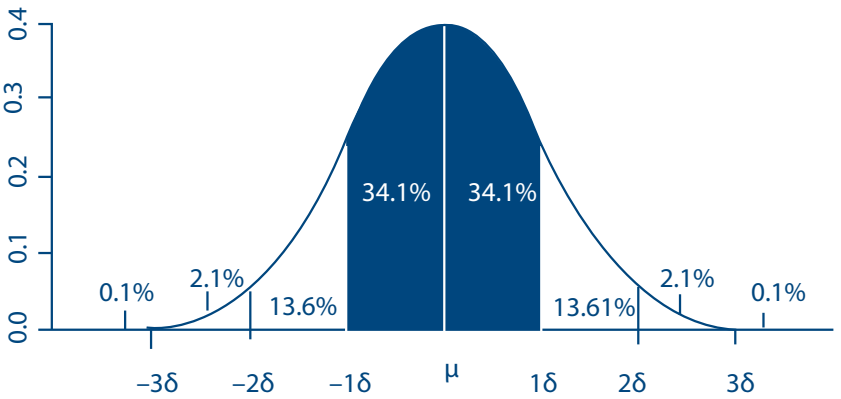


Figura 2.22

Fuente: http: //commons.wikimedia.org/wiki/File:Standard_deviation_diagram_(decimal_comma).svg

Para ejemplificar lo antes descrito, considérese el siguiente problema resuelto.

Problema resuelto

Un almacén realiza en promedio 500 entregas a la semana, más/menos 50. El administrador del negocio desea saber la probabilidad de:

a) Realizar 440 entregas a la semana.
b) Realizar 550 entregas a la semana.
c) Realizar entre 440 y 550 entregas a la semana.

Solución

a) En este caso, primero se calcula el valor de *Z*. Donde:

$\mu = 500$
$\sigma = 50$

$$Z = \left| \frac{440 - 500}{50} \right| = 1.2$$

Enseguida, se ubica la probabilidad en tablas.

$$P(Z = 1.2) = 0.3849 = 38.49\%$$

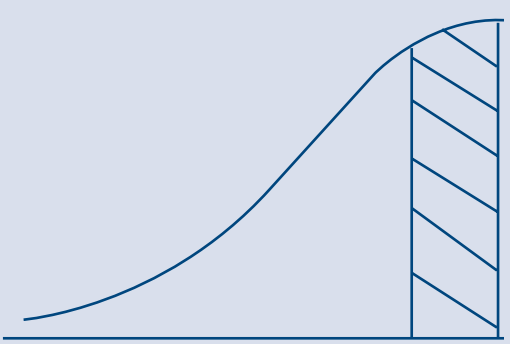


Figura 2.23

Estadísticos e-Books & Papers

Solución *(continuación)*

b) En este caso, primero se calcula el valor de *Z*:

$$Z = \left| \frac{550 - 500}{50} \right| = 1.0$$

Luego, en tablas se ubica la probabilidad:

$$P(Z = 1.0) = 0.3413 = 34.13\%$$

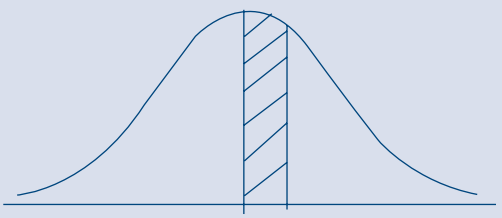


Figura 2.24

c) En este caso, la probabilidad es la suma de las áreas. Por tanto:

$$38.49\% + 34.13\% = 72.62\%$$

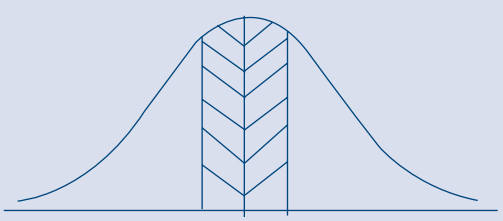


Figura 2.25

2.20 Teorema del límite central

El teorema del límite central tiene como función señalar que las variables aleatorias tienden a distribuirse como la normal; en otras palabras, señala que la distribución normal puede usarse como aproximación a otras funciones de distribución. Por ejemplo:

- La distribución binomial con parámetros *n* y *p* se aproxima a la normal para grandes valores de *n* y *p*, no demasiado cercano a 1 o 0. De tal manera que la normal aproximada tiene los siguientes parámetros: $\mu = np$, $\sigma^2 = np(1 - p)$.
- La distribución de Poisson con parámetro λ se aproxima a la normal para grandes valores de λ. Por lo que la distribución normal aproximada tiene parámetros $\mu = \sigma^2 = \lambda$.

Alerta

El teorema del límite central expone que algunas distribuciones de probabilidad tienden a la normal si se cumplen ciertos valores en sus parámetros.

Estadísticos e-Books & Papers

2.1 Desarrolla el espacio muestral correspondiente al lanzamiento de dos dados.

2.2 Desarrolla el espacio muestral de dos monedas lanzadas simultáneamente.

2.3 En un estudio sobre preferencias de lectura de diarios, se realizó una encuesta a 45 lectores, en la cual se obtuvieron los siguientes resultados:

a) 29 señalaron leer el diario *El Planeta*.

b) 25 dijeron leer *El Mercurio*.

c) 30 mencionaron leer *Atlas*.

Sin embargo, a su vez, 10 lectores señalaron leer *El Planeta* y *El Mercurio*; en tanto, 20 afirmaron leer *El Planeta* y *Atlas*; mientras que 18 dijeron leer *El Mercurio* y *Atlas*; por último, 5 señalaron leer los tres.

Desarrolla el diagrama de Venn correspondiente a la información obtenida de la encuesta.

2.4 En una clase en la que todos los alumnos practican algún deporte, 60% de los estudiantes juega futbol o baloncesto, de los cuales 10% practica ambos deportes y 40% solo juega baloncesto. Determina: ¿cuál será la probabilidad de que, escogido al azar, un alumno de la clase:

a) juegue solo futbol?

b) juegue solo baloncesto?

c) practique uno solo de los deportes?

d) no juegue ni futbol ni baloncesto?

2.5 Un club de lectura de estudiantes de la carrera de Ingeniería aeronáutica recibe tres revistas: *Alas* (A), *Aeronaves* (B) y *Aviones* (C). El número de miembros del club es de 100 estudiantes, cuya preferencia de lectura es la siguiente:

a) 66 leen *Alas*.

b) 38 leen *Aeronaves*.

c) 59 leen *Aviones*.

d) 17 leen *Alas* y *Aeronaves*.

e) 43 leen *Alas* y *Aviones*.

f) 13 leen *Aeronaves* y *Aviones*.

El tutor y el presidente del club están interesados en saber, si se selecciona a un miembro del club al azar, cuál es la probabilidad de que…

a) Lea las tres revistas.

b) Lea *Alas* pero no *Aviones*.

c) Lea *Alas* o *Aviones*, pero no *Aeronaves*.

d) Lea *Alas*, pero no *Aviones* ni *Aeronaves*.

2.6 El departamento de Ciencias Sociales de una universidad cuenta con 800 estudiantes, por lo que decidió realizar un estudio sobre el número de estudiantes que durante el actual semestre cursarán las asignaturas de Metodología de la investigación, Administración y Estadística. A través de una encuesta, se obtuvieron los siguientes resultados:

Tabla 2.5

Metodología	490	**Metodología y Administración**	90
Administración	160	**Metodología y Estadística**	22
Estadística	320	**Administración y Estadística**	78

Con base en esta información y si se selecciona a un alumno al azar, calcula las siguientes probabilidades:

a) De que estudie las tres asignaturas.

b) De que estudie solo Estadística.

c) De que estudie Metodología y Administración.

d) De que estudie Administración y Estadística.

2.7 Como es sabido, un dominó cuenta con 28 fichas, ordenadas en siete fichas por cada serie de números: 0, 1, 2, 3, 4, 5, 6. Halla la probabilidad de que al levantar las fichas de dominó se obtenga:

a) Un número de puntos mayor a 25 puntos.

b) Un número de puntos que sea múltiplo de 6.

c) Un número de puntos mayor a 20, pero menor a 36.

2.8 En la realización de un experimento se utilizan tres cajas. En la primera hay 3 bolas blancas y 2 bolas rojas; en la segunda caja, hay 4 bolas blancas y 3 bolas verdes, y en la tercera caja hay 5 bolas blancas y una bola amarilla. Si se extraen dos bolas de cada caja sin reemplazo, determina: ¿cuál es la probabilidad de obtener una bola roja, una bola verde y una bola amarilla?

2.9 En la realización de un experimento se utilizan 4 cajas, cada una de las cuales tiene en su interior 3 bolas azules y 5 bolas blancas. Si se extrae una bola de cada caja sin reemplazo, ¿cuál es la probabilidad de obtener dos bolas azules y 2 bolas blancas en cualquier orden?

2.10 Considerando el planteamiento del problema 2.9, determina: ¿cuál es la probabilidad de obtener 3 bolas azules y una bola blanca en cualquier orden?

2.11 Considerando el planteamiento del problema 2.9, si todas las bolas se juntan en una sola caja y se extraen 4 bolas de manera consecutiva, sin reemplazo; establece: ¿cuál es la probabilidad de obtener una bola blanca, 2 bolas azules y una bola blanca, en ese orden?

2.12 Se lanzan dos dados al aire y se suman los puntos obtenidos. Determina:

a) La probabilidad de que el número obtenido sea 7.

b) La probabilidad de que el número obtenido sea par.

c) La probabilidad de que el número obtenido sea múltiplo de tres.

Estadísticos e-Books & Papers

2.13 Para la realización de un experimento se emplean dos monedas. Con la primera de estas, la probabilidad de obtener "sol" es de $\frac{2}{3}$, mientras que con la segunda es de $\frac{5}{8}$. Determina la probabilidad de que al lanzar al aire dos monedas salgan:

a) Dos veces sol.

b) Dos veces águila.

c) Una vez águila y una vez sol.

2.14 En una clase universitaria de ciencias hay inscritos 30 alumnos, de ellos 5 estudian física, 15 matemáticas y 10 biología. De estos mismos estudiantes, 22 son mujeres y el resto son hombres. Determina: ¿cuál es la probabilidad de que al escoger un estudiante al azar para pasar al pizarrón, este sea hombre y estudiante de matemáticas?

SR **2.15** Una fábrica que produce coples para tubería de $\frac{3}{4}$" de diámetro reúne los excedentes de su producción en una caja, para su clasificación en el almacén. De esta manera, en la caja hay 100 coples, tanto de pared gruesa como de pared delgada. De estos, 55 coples son cortos, de los cuales 25 son de pared delgada, mientras que 45 son coples largos, de los cuales 15 son de pared gruesa. Si se selecciona un cople al azar de la caja, establece cuál es la probabilidad de que este sea…

a) Un cople corto o de pared delgada.

b) Un cople largo de pared delgada.

c) Un cople corto.

d) Un cople largo o de pared gruesa.

2.16 Como es sabido, el juego del cubilete es un juego de azar que cuenta con 5 dados, normalmente marcados con palos de la baraja americana. Determina: ¿cuál es la probabilidad de obtener cuatro ases en dos tiradas consecutivas?

2.17 Una empresa dedicada al diseño de juegos de mesa, hoy día se encarga de desarrollar las reglas de uno de sus nuevos productos, proponiendo que el jugador deba lanzar dos dados, uno después del otro obligadamente. El ganador será aquel jugador que acumule más puntos, siempre y cuando la suma de los puntos sea de por lo menos 7, para que cuente la tirada. Determina lo siguiente:

a) ¿Cuál es la probabilidad de que la suma de los dados sume al menos 7?

b) ¿Cuál es la probabilidad de que la suma de los dados sea al menos 7, pero que en el segundo dado aparezca el número 4?

c) ¿Cuál es la probabilidad de que la suma de los dados sea al menos 7 y que en el primer dado aparezca el número 2?

2.18 Durante la época de invierno, en un aeropuerto ubicado en la frontera norte del país, la probabilidad de que un aterrizaje se realice bajo lluvia es de 40%, que aterrice bajo nieve es de 25%, y de que aterrice bajo clima despejado es de 35%. En atención a las probabilidades climáticas, los riesgos de despiste son de: 15%, 15% y 7%, respectivamente. Los analistas del aeropuerto están interesados en conocer las siguientes probabilidades:

a) Que un aterrizaje se desarrolle bajo condiciones de riesgo.

b) Que un aterrizaje se desarrolle libre de riesgo.

c) Que un aterrizaje se desarrolle con riesgo bajo lluvia.

d) Que un aterrizaje se desarrolle libre de riesgo bajo nieve.

SR **2.19** En la actualidad, una cadena de restaurantes de comida rápida cuenta con cuatro sucursales en la ciudad y zona metropolitana. Debido a sus pretensiones de extensión, está interesada en captar la preferencia del consumidor con base en la calidad en el servicio. Se sabe que el nivel de concurrencia de sus clientes es de 30%, 24%, 28% y 16%, respectivamente en cada sucursal. Como resultado de su evaluación del SERVQUAL (Service Quality Evaluation) se sabe que 84%, 87%, 80% y 92% de los clientes de cada sucursal, respectivamente, están de acuerdo con el servicio. La dirección de la cadena está interesada en conocer la probabilidad de que la opinión de un cliente escogido al azar sea…

a) Que no esté de acuerdo con el servicio y que sea cliente de la sucursal "1".

b) Que esté de acuerdo con el servicio y que sea cliente de la sucursal "3".

c) Que no esté de acuerdo y que sea cliente de la sucursal "4".

2.20 En una oficina, 70% de los empleados son originarios del Distrito Federal. De ellos, 50% son hombres. En tanto, de los empleados que son originarios de otras partes del país, solo 20% son hombres. Con base en estos datos:

a) Calcula la probabilidad de que un empleado no capitalino sea mujer.

b) Calcula la probabilidad de que un empleado de la oficina sea mujer.

SR **2.21** En un supermercado, 65% de las compras son realizadas por mujeres. De las compras realizadas por estas, 80% supera los 200 pesos, mientras que de las compras realizadas por hombres solo 30% supera esa cantidad. Elegido un comprobante de compra al azar, determina:

a) ¿Cuál es la probabilidad de que este supere los 200 pesos?

b) Si se sabe que el comprobante de compra no supera los 200 pesos, ¿cuál es la probabilidad de que la compra haya sido hecha por una mujer?

SR **2.22** El equipo directivo de una empresa perteneciente al sector hotelero está constituido por 25 personas. Del número total de empleados directivos, 60% son mujeres. De ellos, el gerente tiene que seleccionar a una persona de dicho equipo, para que represente a la empresa en un certamen internacional. Para la selección, decide lanzar una moneda al aire. Si sale sol, selecciona a una mujer y si sale águila, elige a un hom-

Estadísticos e-Books & Papers

bre. Sabiendo que 5 mujeres y 3 hombres del equipo directivo no hablan inglés, determina:

a) La probabilidad de que la persona seleccionada hable inglés.

b) La probabilidad de que la persona seleccionada no hable inglés.

2.23 Una empresa fabricante de bandas para motores de equipo de construcción cuenta con cuatro plantas de producción en el país (A, B, C, D). La producción total (en las cuatro plantas) de bandas para bulldozer es de 5 000 unidades al mes. De la producción total, cada planta manufactura 1 000, 1 150, 1 500 y 1 350 unidades, respectivamente. Se sabe que el porcentaje de unidades defectuosas generadas por cada planta es de 15%, 18%, 12% y 16%, respectivamente. Considerando la información expuesta antes, la gerencia de producción está interesada en conocer, en caso de que se seleccione una banda al azar, cuál es la probabilidad de que…

a) Sea manufacturada sin defectos.

b) Sea manufacturada por la planta D sin defectos.

c) Sea manufacturada por la planta B con defectos.

2.24 En el proceso de producción, una fábrica utiliza líneas de producción. La primera línea produce 30% de la producción total; la segunda línea produce 50% y la tercera línea contribuye con 20%. La tabla siguiente muestra las probabilidades de que se seleccionen productos sin defectos de cada una de las líneas.

Línea 1	0.8
Línea 2	0.7
Línea 3	0.9

Con base en la información anterior, el gerente de control de calidad desea conocer las siguientes probabilidades:

a) Seleccionar un artículo defectuoso de la tercera línea.

b) Seleccionar un artículo de calidad de la segunda línea.

c) Seleccionar un artículo defectuoso de la primera línea.

d) Seleccionar un artículo de calidad de cualquier línea.

2.25 Una empresa de servicios por cable determinó que la probabilidad de contratación de los servicios en una zona de clase media de reciente creación es la siguiente:

Cliente	Probabilidad de contratación
Casa	65%
Departamento	35%

De acuerdo con los resultados del estudio, estima la probabilidad de contratación de los servicios, como se presentan a continuación:

Servicio	Probabilidad de servicio
Teléfono	50%
TV	30%
Internet	20%

De acuerdo con los datos anteriores, determina lo siguiente:

a) La probabilidad de que se contraten servicios de telefonía.

b) La probabilidad de que un contrato de Internet sea de un departamento.

c) La probabilidad de que un contrato de televisión sea de una casa.

d) La probabilidad de que un contrato sea de telefonía, televisión e Internet.

2.26 La compañía de juguetes responsable de la fabricación de la muñeca de moda, hace poco tiempo lanzó al mercado otro kit de guardarropa. Este incluye: cuatro vestidos (azul, amarillo, rojo y blanco), tres pares de zapatos (rojo, blanco y negro), tres bolsas (blanca, roja y rosa). Desarrolla el árbol de decisión, con el fin de determinar las diversas formas de combinación del guardarropa.

2.27 Una cadena de hamburguesas tomó la decisión de rediseñar la composición de sus productos. En esta nueva etapa, el cliente puede escoger tres de los siguientes ingredientes extras:

a) Tocino

b) Queso

c) Salsa especial

d) Lechuga y pepinillos

e) Aderezo agridulce

Con base en estos datos, desarrolla el árbol de decisión, para determinar el número de formas en que se pueden combinar los ingredientes extras.

2.28 El estudio demográfico de un país arrojó los siguientes datos:

a) Tres de cada siete nacimientos son varones.

b) En promedio, las parejas procrean cuatro hijos.

Desarrolla el árbol de decisión, para determinar las posibles combinaciones de los hijos de cada pareja.

2.29 Con base en los datos del problema 2.26, aplica el principio de la multiplicación para confirmar el número de formas posibles de combinar el nuevo guardarropa de la muñeca de moda.

2.30 De acuerdo con los datos del problema 2.27, la gerencia de servicio está interesada en determinar el número de combinaciones posibles de hamburguesas que los clientes pueden formar.

2.31 Una agencia de automóviles desarrolla una estrategia de promoción para vender todos los autos modelo 2010, que aún se hallan en la agencia, la cual consiste en ofrecer un paquete que consiste en escoger tres accesorios, de cuatro posibles, a un precio especial, los cuales son:

Estadísticos e-Books & Papers

Accesorio	Descripción
A	Rines deportivos
B	Manos libres para celular en el tablero
C	Quemacocos
D	Equipo de sonido para iPod integrado

El gerente de servicio desea saber cuántas y cuáles serían las combinaciones de los paquetes. Determina todas las probabilidades.

SR **2.32** Una fábrica de alarmas para autos establece que la probabilidad de que una de sus alarmas falle es de 0.005. Si cada uno de los lotes de producción consta de 2000 unidades, la gerencia de calidad está interesada en saber las siguientes probabilidades:

a) De que fallen exactamente 10 unidades.

b) De que fallen más de 10 unidades.

c) De que falle 5% de las unidades.

d) De que falle 1% de las unidades

2.33 Con respecto al estudio demográfico, referido en el problema 2.28, los responsables del mismo están interesados en conocer las siguientes probabilidades:

a) De que conciban un varón.

b) De que conciban cuatro varones.

c) De que no conciban ningún varón.

d) De que conciban dos varones y dos mujeres.

2.34 Durante la realización del examen para obtener la licencia de manejo, se conoció que 3 de cada 10 candidatos son rechazados la primera vez que realizan dicho examen. Determina la distribución de probabilidad de que se acepten 5, 6, 7, 8, 9, 10 candidatos.

2.35 La probabilidad de que un artículo producido por una fábrica no sea defectuoso es de 99.80%. Si se envía un cargamento de 10000 artículos a diversos almacenes, hallar la probabilidad de que…

a) en el embarque vayan 10 artículos defectuosos.

b) en el embarque vayan no más de siete artículos defectuosos.

c) en el embarque vayan entre cinco y 10 artículos defectuosos.

2.36 Una fábrica de equipos de cómputo dio a conocer que de cada 20000 discos duros que produce, 19 880 discos cumplen con las especificaciones de control de calidad. Si los lotes de embarque están compuestos por 600 unidades, la gerencia de control de calidad quiere conocer las siguientes probabilidades:

a) De que en un lote vayan no más de 6 discos duros defectuosos.

b) De que en un lote vayan más de 6 discos duros defectuosos.

c) De que en un lote vayan más de 6, pero menos de 10 discos duros defectuosos.

SR **2.37** Las estadísticas de salud más recientes de nuestro país, afirman que en la zona sur-oriente de la Ciudad de México se presenta una alta incidencia de malestares estomacales, provocados por la falta de higiene (esto es, 200 casos por cada 10000 habitantes). Supón que también se realizan exámenes a 1000 habitantes de un municipio conurbado y se asume que para estos la tasa de incidencia es la misma que para toda la región sur-oriente de la Ciudad de México. Determina:

a) ¿Cuál es la probabilidad de que ninguna de las personas examinadas padezca alguna infección estomacal?

b) ¿Cuál es la probabilidad de que al menos tres personas padezcan alguna infección estomacal?

c) ¿Cuál es la probabilidad de que al menos ocho personas padezcan alguna infección estomacal?

2.38 Una empresa dedicada a la fabricación de focos incandescentes, reporta que 7% de la producción es defectuosa. Halla la probabilidad de que en una muestra de 270 bombillas suceda que…

a) más de cinco focos sean defectuosos.

b) entre uno y tres focos sean defectuosos.

c) dos focos o menos sean defectuosos.

2.39 Después de un estudio realizado por una institución bancaria, se determinó que existe una probabilidad de 40% de que los clientes que cuentan con tarjeta de crédito paguen antes de su fecha de corte. Para revisar la vigencia de esta condición, el gerente de crédito selecciona cinco cuentas, con el fin de conocer: ¿cuál es la probabilidad de que se presenten cargos por intereses en 0, 1, 2, 3, 4 y 5 cuentas? Con base en ese objetivo desarrolla la curva de distribución discreta correspondiente.

2.40 En la inspección de hojalata producida por un proceso electrolítico continuo, se identifican 0.2 imperfecciones en promedio por minuto. Determina las probabilidades de identificar:

a) Una imperfección en tres minutos.

b) Dos o menos imperfecciones en cinco minutos.

c) Una imperfección o menos en 15 minutos.

2.41 De acuerdo con el planteamiento del problema 2.36, resuelve este por aproximación de Poisson a la binomial.

2.42 Considerando el planteamiento del problema 2.37, resuelve por aproximación de Poisson a la binomial.

2.43 De acuerdo con la información expuesta en el problema 2.34, acerca del examen de manejo, se sabe que los examinadores aplican el examen a 2.5 candidatos por hora en promedio. Si la jornada de trabajo de cada empleado es de 8 horas, ¿cuál es la probabilidad de que durante una jornada de trabajo se examinen 20, 21, 22, 23, 24, 25 candidatos?

Estadísticos e-Books & Papers

Problemas para resolver

2.44 Una empresa dedicada a la fabricación de placas para pisos de madera, encontró que después del proceso de barnizado existe 0.2% de probabilidad de hallar un defecto por cada metro cuadrado, extensión donde el límite de imperfecciones máximo es de tres por metro cuadrado. Ante esta situación, la gerencia de control de calidad desea conocer lo siguiente:

a) La probabilidad de encontrar 3 defectos por metro cuadrado.

b) La probabilidad de encontrar no más de 3 defectos por metro cuadrado.

c) La probabilidad de encontrar más de 2 defectos por metro cuadrado.

2.45 La probabilidad de acertar a un blanco en movimiento con una nueva metralleta montada en un auto blindado maniobrando sobre terreno escabroso es de 0.004% en cada disparo. Si se realizan 3000 disparos y se considera un análisis por distribución binomial, ¿cuál es la probabilidad de acertar al menos cuatro disparos?

2.46 La probabilidad de que un individuo sufra una reacción negativa como consecuencia de la aplicación de un suero es de 0.001%. A últimas fechas, el suero ha sido aplicado a 2000 personas. Determinar mediante la distribución de Bernoulli lo siguiente:

a) La probabilidad de que se presente la reacción exactamente en tres personas.

b) La probabilidad de que se presente la reacción en al menos tres personas.

c) La probabilidad de que se presente la reacción en más de tres personas.

2.47 Con el objetivo de solucionar en la medida de lo posible el problema de obesidad entre los estudiantes de una preparatoria, la dirección de servicios médicos del plantel encontró que la media de los pesos de 500 estudiantes fue de 68.5 kilogramos, considerando una tolerancia de 7 kilogramos. La dirección de servicios médicos desea conocer lo siguiente:

a) La probabilidad de que los estudiantes tengan un peso entre 54.5 y 70.0 kilogramos.

b) La probabilidad de hallar estudiantes que tengan un peso por arriba de los 84 kilogramos.

c) La probabilidad de hallar estudiantes que pesen menos de 58 kilogramos.

d) La probabilidad de ubicar estudiantes que pesen entre 54.5 y 58 kilogramos, ya que es el rango que corresponde a su edad y desarrollo físico.

2.48 Un proyecto de ingeniería tiene un presupuesto promedio mensual de 8300 dólares, con una variación de 764 dólares. De acuerdo con estos datos, establecer cuál es la probabilidad de que el presupuesto…

a) Se encuentre entre 8300 dólares, más-menos 780 dólares.

b) Se encuentre entre 9000 y 11000 dólares.

c) Sea menor a 6873 dólares.

2.49 De acuerdo con un estudio que contempló la revisión de las cuentas de ahorro populares, se estableció que el monto promedio mensual de cada cuenta es de 830 pesos, presentando una variación de 76.40 pesos. Determina: ¿cuál es la probabilidad de que el monto de una cuenta de ahorro se ubique entre…

a) 830 pesos más-menos 80 pesos?

b) 900 y 1100 pesos?

c) sea menor a 700 pesos?

2.50 Considerando la función de densidad de la distribución normal, grafica en Excel $\mu = 500$ y $\sigma = 460$ en un rango entre -500 y 2500.

PROBLEMAS RETO

1 Desarrolla el espacio muestral del lanzamiento de tres dados, de tal manera que se establezcan las siguientes probabilidades:

a) Que la suma de puntos sea mayor a 9.

b) Que la suma de puntos sea entre 15 y 18.

c) Que la suma de puntos sea exactamente 12.

2 Considerando la función de densidad de la distribución normal, realiza la tabla de áreas bajo la curva comprendida entre $Z = 0$ y $Z = 3$, con una precisión de 0.01 unidades.

Estadísticos e-Books & Papers

REFERENCIAS

Levin, Richard I. y Charles A. Kirkpatrick (1989). *Enfoques cuantitativos a la administración*. México: Continental y McGraw-Hill.

Lind, Douglas A., William G. Marchal y Samuel A. Wathen (2005). *Estadística aplicada a los negocios y a la economía* (12a. ed.). México: McGraw-Hill.

Maris, Diez Stella (2005). *Estadística aplicada a los negocios utilizando Microsoft EXCEL*. Argentina: MP Ediciones.

Quevedo, Urias Héctor y Blanca Rosa Pérez Salvador (2008). *Estadística para Ingeniería y Ciencias*. México: Grupo Editorial Patria.

DIRECCIONES ELECTRÓNICAS

Problemas resueltos de probabilidad (nivel básico)

[http://www.vitutor.com/pro/2/a_g.html]

Problemas de probabilidad resueltos y propuestos

[http://webdelprofesor.ula.ve/ciencias/jlchacon/materias/discreta/probpro.pdf]

Simulador de la distribución binomial

[http://virtual.uptc.edu.co/ova/estadistica/applets/simulacionbino/Simulacion.htm]

Portal académico CCH-distribuciones de probabilidad

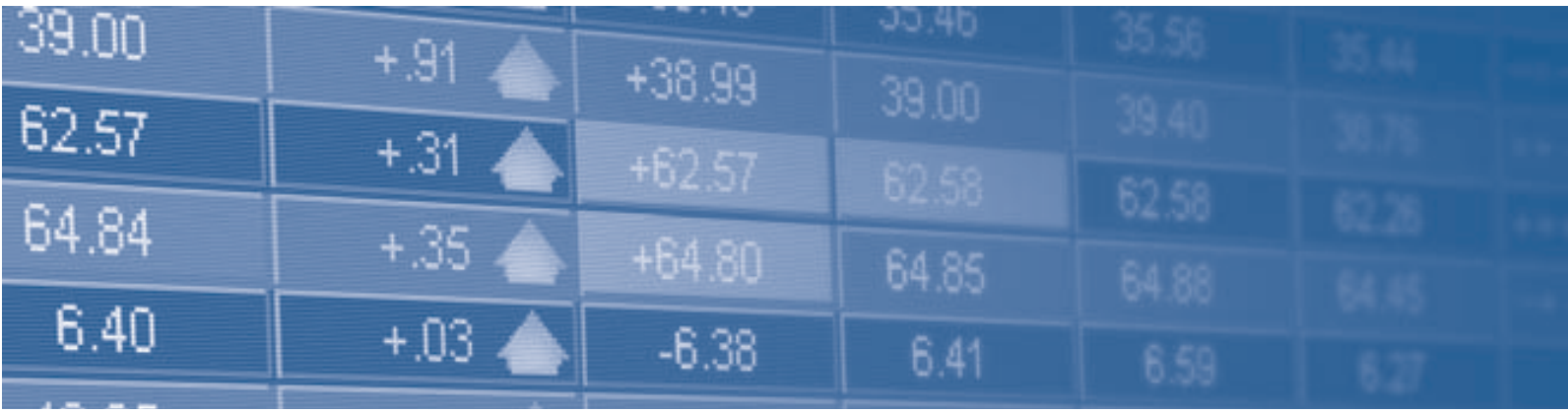
[http://portalacademico.cch.unam.mx/alumno/sitiosdeinteres/matematicas/estadistica2]

Distribución normal

[http: //es.wikipedia.org/wiki/Distribuci%C3%B3n_normal]

Estadísticos e-Books & Papers

Estadísticos e-Books & Papers



Estadística inferencial

OBJETIVOS

- Comprender la importancia de la estadística inferencial.
- Entender los conceptos de población y muestra, así como establecer la diferencia entre ellos.
- Distinguir la diferencia entre muestras probabilísticas y no probabilísticas.
- Entender los conceptos de estadístico y parámetro, así como la diferencia entre ellos.
- Deducir las fórmulas para determinación del tamaño muestral.
- Comprender el concepto de error muestral.
- Comprender el concepto de error estándar.
- Entender el concepto de estimador.
- Distinguir la diferencia entre un estimador sesgado y uno insesgado.
- Entender el concepto de nivel de confianza.
- Comprender la trascendencia del tamaño muestral en relación con el análisis inferencial.
- Entender el concepto de grados de libertad.
- Comprender las características de la distribución *t* de Student.

¿QUÉ SABES?

- ¿Cuál es la trascendencia de la estadística inferencial?
- ¿Cuál es la diferencia entre una población y una muestra?
- ¿Cuál es la diferencia entre las muestras probabilísticas y no probabilísticas?
- ¿Cuál es la diferencia entre un estadístico y un parámetro?
- ¿Cómo se puede determinar el tamaño muestral?
- ¿Qué es el error muestral?
- ¿Qué es el error estándar?
- ¿Qué es un estimador?
- ¿Cuál es la diferencia entre un estimador sesgado y uno insesgado?
- ¿Qué es el nivel de confianza?
- ¿Qué importancia tiene el tamaño muestral para el desarrollo del análisis inferencial?
- ¿Qué son los grados de libertad?
- ¿Cuáles son las principales características de la distribución *t* de Student?

Estadísticos e-Books & Papers

3.1 Introducción

Es común que en la vida diaria se hable de muestras y de resultados de estudios estadísticos relativos a opiniones sobre un tema o referentes a las características de un producto o servicio, por lo que es de interés exponer la teoría y los conceptos básicos que permiten estructurar los procesos de análisis correspondientes.

3.2 Concepto y propósito de la estadística inferencial

Alerta

La estadística inferencial expone el estudio del comportamiento de un fenómeno que ocurre en una población a través del estudio y análisis de muestras.

La estadística inferencial expone el estudio del comportamiento de un fenómeno que ocurre en una población a través del estudio y análisis de muestras; de manera práctica, es establecer las dimensiones estadísticas de una población (parámetros) a través de las dimensiones estadísticas de las muestras (estadísticos).

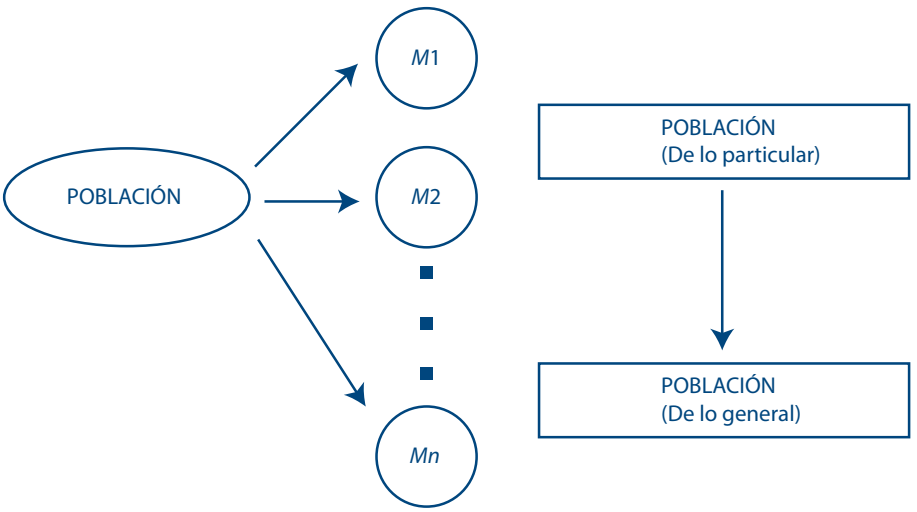


Figura 3.1

Alerta

Las dimensiones estadísticas de una muestra se denominan estadísticos, mientras que en la población se denominan parámetros.

Las dimensiones estadísticas son las mismas en cuanto a su concepto; sin embargo, operacional y simbólicamente se distinguen como se muestra en la tabla 3.1.

Tabla 3.1 Dimensiones estadísticas

Dimensión estadística	Población (parámetros)	Muestra (estadísticos)
Media	μ	x
Desviación estándar	σ	S
Número de elementos	N	n

Concepto de población

Alerta

Una población es el conjunto de personas, objetos o entidades que cumplen con ciertas características definidas por el observador.

Se define **población** como el conjunto de personas, objetos o entidades que cumplen con ciertas características definidas por el observador.

Las poblaciones se pueden clasificar como:

Finitas, cuando el número de elementos es contabilizable y ubicable.

Infinitas, cuando el número de elementos es tan grande que se dificulta su contabilización pese a ser ubicables.

Difusas, cuando el número de elementos es difícilmente contabilizable debido a que la ubicación de los elementos no se puede establecer.

Estadísticos e-Books & Papers

Concepto de muestra

Se define como muestra al **conjunto** de elementos representativo de una población.

Las muestras se pueden clasificar en:

I. **Probabilísticas**, donde el proceso de selección de los elementos que la estructuran se basa en las reglas de la probabilidad. Entre las muestras probabilísticas se pueden citar las siguientes:

- **Simples.** Son aquellas donde los elementos de una población son etiquetados numéricamente de manera que puedan ser seleccionados en forma aleatoria, y pueden ser con o sin reemplazo.
- **Sistemáticas.** Son aquellas donde los elementos de la población son etiquetados numéricamente y catalogados por intervalos, seleccionándose en forma aleatoria de cada intervalo y repitiéndose el procedimiento hasta completar la muestra.
- **Estratificadas.** Son aquellas donde los elementos de la población se clasifican en subgrupos o estratos de acuerdo con una característica definida, procediendo a realizar un muestreo simple de cada estrato, de manera que la muestra contenga elementos de todos los estratos.
- **De conglomerado.** Son aquellas donde los elementos de la población se catalogan en subgrupos de acuerdo con una característica de orden general, como lo es una clasificación por género, colonia de residencia o lugar de nacimiento. Así, una vez formados los conglomerados, estos deben ser seleccionados en forma aleatoria, de manera que se elijan tantos elementos de cada uno hasta que se complete la muestra.

II. **No probabilísticas**, donde los elementos de una población no cuentan con una probabilidad de ser seleccionados para estructurar una muestra. Dentro de las muestras no probabilísticas se pueden citar las siguientes:

- **De conveniencia.** Son aquellas donde los elementos son seleccionados por cumplir con ciertas condiciones o características que faciliten el análisis del fenómeno; inclusive pueden estructurarse por voluntarios.
- **De criterio.** Son aquellas donde los elementos son seleccionados con base en un criterio o juicio del observador.
- **Por cuotas.** Son aquellas donde la población se cataloga en estratos, y se procede a seleccionar tantos elementos de cada uno hasta que se complete la muestra.
- **De bola de nieve.** Son aquellas donde se selecciona un cierto número de elementos a los cuales se les solicita ofrezcan referencias sobre otros elementos de la población, y se repite el procedimiento hasta completar la muestra o satisfacer algún requisito en el número de elementos.

Alerta

Una muestra es el conjunto de elementos representativo de una población, y puede ser probabilística o no probabilística.

3.3 Estimadores

Como se ha expuesto, el propósito del análisis inferencial es determinar el valor de los parámetros de la población a través de los estadísticos de las muestras. Por consiguiente, se puede señalar que cuando el valor del estadístico coincide con el valor del parámetro se dice que el estadístico es un estimador insesgado, o sea que no existe diferencia entre el valor del parámetro y el estadístico; pero cuando no coincide, y existe diferencia, se dice que el estadístico es un estimador sesgado.

De hecho, se puede señalar que si se pudieran analizar todas las muestras, consistentes y válidas, de una población, se pueden estimar sus parámetros.

Para demostrar lo expuesto considérese el siguiente problema.

Alerta

Un estadístico es un estimador insesgado cuando no existe diferencia con el valor del parámetro y el estadístico; y cuando existe diferencia es un estimador sesgado.

Problema resuelto

A partir de la población,

$$P = \{10, 17, 24, 32\}$$

Determina todas las muestras, consistentes y válidas, de tres elementos a efecto de calcular y comparar sus estadísticos con los parámetros de la población.

Estadísticos e-Books & Papers

Alerta

Si pudieran analizarse todas las muestras disponibles de una población, se estaría analizando a la población misma.

Solución

En primer lugar se determina el número de muestras, de manera que estas no se repitan (consistentes), por lo que se aplica la fórmula de las combinaciones:

$$ {}_4C_3 = \frac{4!}{3!(4-3)!} = 4 $$

Por lo que, desarrollando los arreglos, las muestras resultantes son las siguientes:

m_1 = 10, 17, 24 $\qquad$ m_2 = 10, 17, 32

m_3 = 10, 24, 32 $\qquad$ m_4 = 17, 32, 24

Se procede a calcular el valor de la media y varianza de cada muestra.

m_1 $\qquad \overline{X}$ = 17.00

X	$(x-\overline{X}_1)^2$
10	49.00
17	0.00
24	49.00
Σ =	98.00

$$ S_1^2 = \frac{\Sigma_{i=1}^{n}(x_i - \overline{X}_1)^2}{n} = \frac{98.00}{3} = 32.66 $$

m_2 $\qquad \overline{X}_2$ = 19.67

X	$(x-\overline{X}_2)^2$
10	93.3156
17	7.07556
32	152.2756
Σ =	252.67

$$ S_2^2 = \frac{252.67}{3} = 84.22 $$

m_3 $\qquad \overline{X}_3$ = 22.00

X	$(x-\overline{X}_3)^2$
10	144.00
24	4.00
32	100.00
Σ =	248.00

$$ S_3^2 = \frac{248.00}{3} = 82.66 $$

m_4 $\qquad \overline{X}_4$ = 24.33

X	$(x-\overline{X}_4)^2$
17	53.72
32	58.82
24	0.108
Σ =	112.64

$$ S_4^2 = \frac{112.64}{3} = 37.54 $$

Se obtiene el valor de la media y la variación en la población.

μ = 20.75

X	$(x-\mu)^2$
10	115.5625
17	14.0625
24	10.5625
32	126.5625
Σ =	266.75

$$ \sigma^2 = \frac{266.75}{4} = 66.68 $$

Estadísticos e-Books & Papers

Solución *(continuación)*

El valor representativo de las medias muestrales es el promedio de las mismas, por lo que se calcula la media de medias y se compara con la media de la población.

$$\overline{\overline{X}} = \frac{17 + 19.67 + 22 + 24.33}{4} = 20.75$$

$$\mu = 20.75$$

$$\overline{\overline{X}} = \mu$$

Por tanto, se puede concluir que la media de medias es igual a la media poblacional, o sea, la media de las medias es el estimador insesgado de la media poblacional.

De igual forma se obtiene el promedio de las varianzas muestrales.

$$\overline{S}^2 = \frac{32.66 + 84.22 + 82.66 + 37.54}{4} = 59.27$$

Por lo que se puede apreciar que, a diferencia de los promedios de las medias, el valor promedio de las varianzas muestrales es diferente a la varianza poblacional.

$$\sigma^2 \neq \overline{S}^2$$

La pregunta lógica ante este resultado es: ¿por qué ocurre esto? Lo anterior se explica en razón de que tanto la población como las muestras como conjuntos pueden dar origen a otros tantos subconjuntos, pero de hecho no pueden variar en sí mismos, por lo que las varianzas deben calcularse considerando $N - 1$ y $n - 1$.

Por tanto, se calculan las varianzas muestrales y poblacional, y se procede a comparar el valor promedio de las varianzas muestrales con la varianza poblacional.

$$\hat{S}^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}$$

$$\hat{S}_1^2 = 49.00 \qquad \hat{S}_2^2 = 126.34 \qquad \hat{S}_3^2 = 124.00 \qquad \hat{S}_4^2 = 56.32$$

$$\hat{S}_{Prom}^2 = \frac{49.00 + 126.34 + 124.00 + 56.32}{4} = 88.91$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N-1} = 88.91$$

Por tanto se concluye que $\hat{S}_{Prom}^2$ es un estimador insesgado de σ^2.

3.4 Distribución de las medias muestrales

Si se pudiera contar con todas las muestras, consistentes y válidas, de una población, se podría observar que las medias muestrales también cuentan con una distribución de frecuencias; esto es, se puede observar cuántas medias se repiten.

Para demostrarlo, considérese el siguiente problema.

Problema resuelto

A partir de la población que se expone construye la distribución de las medias muestrales a partir de muestras de 4 elementos:

$$P = \{3, 1, 2, 4, 3, 5\}$$

Estadísticos e-Books & Papers

Solución

Con base en el número de elementos de la población ($n = 6$) se determina el número de muestras de 4 elementos:

$$P = \{3, 1, 2, 4, 3, 5\}$$

$$_6C_4 = \frac{6!}{4!(6-4)!} = 15$$

Para determinar las muestras se procede a etiquetar cada uno de los elementos de la población:

3	1	2	4	3	5
a	b	c	d	e	f

Por lo que las muestras correspondientes son las siguientes:

										$\overline{X}$
m_1	a	b	c	d	3	1	2	4	2.50
m_2	a	b	c	e	3	1	2	3	2.25
m_3	a	b	c	f	3	1	2	5	2.75
m_4	a	b	d	e	3	1	4	3	2.75
m_5	a	b	d	f	3	1	4	5	3.25
m_6	a	b	e	f	3	1	3	5	3.00
m_7	a	c	d	e	3	2	4	3	3.00
m_8	a	c	d	f	3	2	4	5	3.50
m_9	a	c	e	f	3	2	3	5	3.25
m_{10}	a	d	e	f	3	4	3	5	3.75
m_{11}	b	c	d	e	1	2	4	3	2.50
m_{12}	b	c	d	f	1	2	4	5	3.00
m_{13}	b	c	e	f	1	2	3	5	2.75
m_{14}	b	d	e	f	1	4	3	5	3.25
m_{15}	c	d	e	f	2	4	3	5	3.50

Donde el valor de la media de medias es: $\overline{\overline{X}} = 3.00$.

Asimismo, el valor de la media poblacional es: $\mu = 3.00$.

Se procede a graficar las frecuencias de las medias de acuerdo con la siguiente tabla:

$\overline{X}$	f
2.25	1
2.50	2
2.75	3
3.00	3
3.25	3
3.50	2
3.75	1
4.00	0
4.25	0

527
Estadísticos e-Books & Papers

Solución *(continuación)*

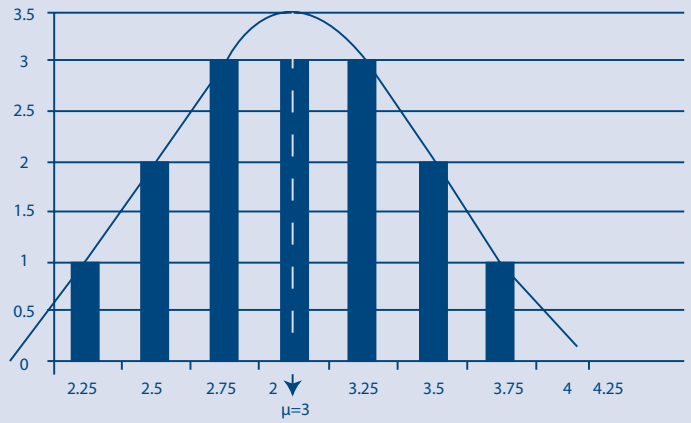


Figura 3.2

Observa la localización de la media poblacional, la cual se ubica en el centro de la distribución.

Para complementar lo expuesto se plantea resolver el siguiente problema con base en las condiciones que se exponen.

Problema resuelto

A partir de la población del problema anterior, construye la distribución de las medias muestrales de 5 elementos.

Solución

A partir del número de elementos de la población (n = 6) se determina el número de muestras de 5 elementos:

$$ {}_6C_5 = \frac{6!}{5!(6-5)!} = 6 $$

Por lo que las muestras correspondientes son las siguientes:

											$\overline{X}$
m_1	a	b	c	d	e	3	1	2	4	3	2.60
m_2	a	b	c	d	f	3	1	2	4	5	3.00
m_3	a	b	c	e	f	3	1	2	3	5	2.80
m_4	a	b	d	e	f	3	1	4	3	5	3.20
m_5	a	c	d	e	f	3	2	4	3	5	3.40
m_6	b	c	d	e	f	1	2	4	3	5	3.00

Donde el valor de la media de medias es: $\overline{X}$ = 3.00.

Asimismo, el valor de la media poblacional es: μ = 3.00.

Se procede a graficar las frecuencias de las medias de acuerdo con la siguiente tabla:

Alerta

Las medias muestrales tienden a distribuirse normalmente, donde la tipificación de la curva depende del número de muestras en razón de su tamaño.

Estadísticos e-Books & Papers

Solución *(continuación)*

$\bar{X}$	f
2.60	1
2.80	1
3.00	2
3.20	1
3.40	1

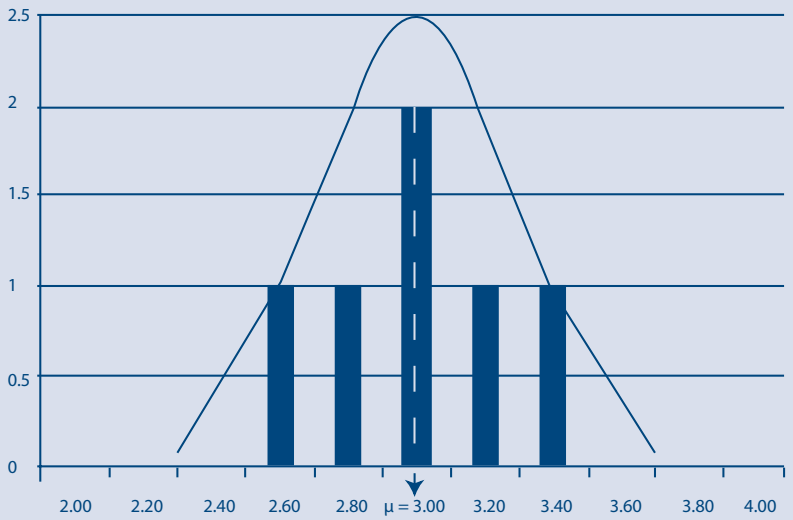


Figura 3.3

En la figura se observa la localización de la media poblacional, la cual se ubica en el centro de la distribución.

Por obvias razones la distribución de las medias dependerá de la estructura y el número de muestras a considerar, ya que si se cambia el número de elementos que estructuran la muestra, también cambiará la forma de la distribución, pero el valor de la media no cambiará.

El error muestral

Alerta

La diferencia entre el estadístico y el parámetro de la media se denomina error muestral de la media.

El concepto expone que puede existir diferencia entre la media muestral (estadístico) y la media poblacional (parámetro) debido a que por ciertas circunstancias no es posible tener acceso a los elementos de la población; en otras palabras, existe una deficiencia en la estructuración de la muestra que genera una diferencia entre el estadístico y el parámetro de la media, donde a la diferencia se le denomina error muestral de la media, el cual matemáticamente queda expresado como:

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Sin embargo, para fines prácticos, si el tamaño de la población N se considera demasiado grande o infinito, el error muestral se puede definir como:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

529
Estadísticos e-Books & Papers

Para ejemplificar lo anterior, se calculan los errores muestrales de los problemas propuestos en la sección 3.3.

Problema resuelto

Considerando los cálculos de la distribución de las medias muestrales de la población:

$$P = \{3, 1, 2, 4, 3, 5\}$$

Para tamaños muestrales $n = 4$ y $n = 5$, calcula los valores de los errores muestrales para cada caso.

Solución

a) Para el caso de $n = 4$ se tiene lo siguiente:

$$N = 6 \qquad n = 4 \qquad \sigma = 1.2909$$

$$\sigma_{\overline{X}} = \frac{\sigma^2}{n} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{1.2909}{\sqrt{4}} \cdot \sqrt{\frac{6-4}{6-1}} = 0.4082$$

b) Para el caso de $n = 5$ se tiene lo siguiente:

$$N = 6 \qquad n = 5 \qquad \sigma = 1.2909$$

$$\sigma_{\overline{X}} = \frac{\sigma^2}{n} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{1.2909}{\sqrt{5}} \cdot \sqrt{\frac{6-5}{6-1}} = 0.2582$$

La diferencia entre los valores de los errores muestrales se puede explicar de manera gráfica, ya que en las gráficas de las distribuciones de las medias se aprecia que cuando $n = 4$ la curva tiene mayor amplitud, condición que no se observa en la curva cuando $n = 5$.

Cálculo de probabilidades de ocurrencia sobre las medias muestrales

Pese a que la distribución de medias muestrales es una variable discreta, se puede observar que su gráfica corresponde a una variable aleatoria continua pudiéndose estimar la probabilidad de seleccionar una muestra con una media muestral con valor $\overline{X}$ cualquiera, por lo que la fórmula de la z se ajusta como sigue:

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

De manera que se pueden determinar las probabilidades utilizando las tablas de áreas bajo la curva normal. Para ejemplificar se muestra el siguiente problema.

Problema resuelto

Considerando los parámetros y estadísticos resultantes de la distribución de medias cuando $n = 4$, ¿cuál es la probabilidad de seleccionar una muestra con una media entre 3.0 y 3.4?

Solución

Los datos necesarios para el cálculo son:

$$\mu = 3 \qquad \sigma_{\overline{X}} = 0.4082 \qquad \overline{X} = 3.4$$

Estadísticos e-Books & Papers

Solución *(continuación)*

por tanto,

$$z = \left| \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \right| = \left| \frac{3.4 - 3}{0.4082} \right| = 0.98$$

Para $P(z = 0.98) = 0.3365 = 33.65\%$

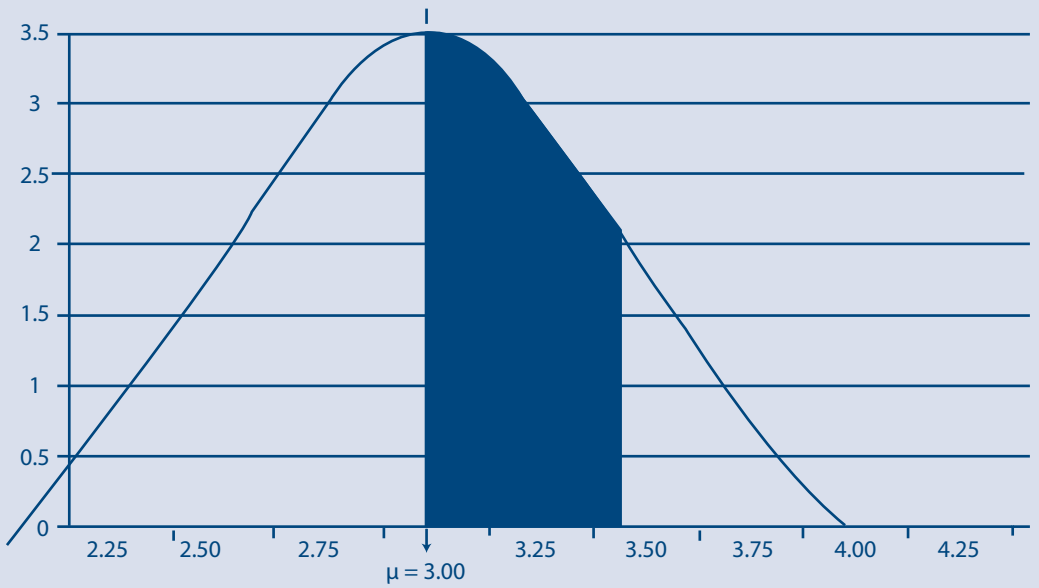


Figura 3.4

Alerta

La distribución normal de las medias muestrales permite calcular la probabilidad de ocurrencia de una muestra con una media que tenga un valor determinado.

3.5 Distribución muestral de las proporciones

De manera similar a la distribución de las medias, puede resultar de interés establecer la relación de la proporción en que ciertos elementos de una población cumplan con ciertas condiciones y la distribución de esta proporción dentro de las muestras; en otras palabras, la distribución muestral de la proporción refiere al conjunto de proporciones de todas las muestras del mismo tamaño, consistentes y válidas, de una población.

Por tanto se puede establecer que la proporción promedio se puede determinar como:

$$\overline{p} = \frac{\Sigma_{i=1}^{Nm} pi}{Nm}$$

Donde:

$\overline{p}$, = Promedio muestral de las proporciones.

pi, = Proporción de la muestra i (i = 1, 2, 3,... Nm)

Nm = Número de muestras.

Asimismo, el error estándar ($\sigma_{\overline{p}}$) de las proporciones queda definido como:

$$\sigma_{\overline{p}} = \sqrt{\frac{PQ}{n} \cdot \frac{N-n}{N-1}}$$

Estadísticos e-Books & Papers

Donde:

P = Valor de la proporción dentro de la población.

Q = Valor de la proporción de que no cumpla dentro de la población ($Q = 1 - P$).

N = Número de elementos de la población (tamaño de la población).

n = Número de elementos de la muestra (tamaño de la muestra).

Asimismo, de manera práctica el error estándar se puede definir como:

$$\sigma_{\bar{p}} = \sqrt{\frac{PQ}{n} \cdot \frac{N-n}{N-1}}$$

En consecuencia, se puede determinar el valor de la probabilidad de que se obtenga una muestra que cumpla con una proporción determinada (p) aplicando la curva de distribución normal, donde la fórmula de conversión a unidades estandarizadas se ajusta como sigue:

$$z = \frac{p - P}{\sigma_{\bar{p}}}$$

Para mostrar lo antes expuesto se plantea el siguiente problema.

Problema resuelto

Una oficina de seguridad y vialidad realiza un estudio sobre el número de conductores multados por infracciones al reglamento de tránsito considerando la siguiente población:

Multados	No multados
A	e
B	f
C	g
D	h

Los responsables del estudio están interesados en realizar un estudio sobre la distribución de la proporción de conductores multados considerando muestras de 5 conductores sobre la población.

Solución

Se observa que la proporción de conductores multados en la población es de 50% ($P = 50\%$).

Donde las proporciones dentro de las muestras de 5 elementos son las que se expresan y siendo la proporción promedio la siguiente:

$$P = 50\% \qquad Q = 1 - P = 50\%$$

$$\overline{p} = \frac{\Sigma_{i=1}^{Nm} pi}{Nm} = \frac{22}{44} = 50\%$$

Donde el valor del error muestral es:

$$\sigma_{\bar{p}} = \sqrt{\frac{PQ}{n} \cdot \frac{N-n}{N-1}} = \sqrt{\frac{0.5 \cdot 0.5}{5} \cdot \frac{6-5}{6-1}} = 0.1$$

Estadísticos e-Books & Papers

Solución *(continuación)*

						p
m_1	a	E	f	g	h	0.2
m_2	b	e	f	g	h	0.2
m_3	c	e	f	g	h	0.2
m_4	d	e	f	g	h	0.2
m_5	a	b	e	F	g	0.4
m_6	a	b	e	F	h	0.4
m_7	a	b	f	g	h	0.4
m_8	a	c	e	F	g	0.4
m_9	a	c	e	F	h	0.4
m_{10}	a	c	f	g	h	0.4
m_{11}	a	d	e	F	g	0.4
m_{12}	a	d	e	F	h	0.4
m_{13}	a	d	f	g	h	0.4
m_{14}	b	c	e	F	g	0.4
m_{15}	b	c	e	F	h	0.4
m_{16}	b	c	f	g	h	0.4
m_{17}	b	d	e	F	g	0.4
m_{18}	b	d	e	F	h	0.4
m_{19}	b	d	f	g	h	0.4
m_{20}	c	d	e	F	g	0.4
m_{21}	c	d	e	F	h	0.4
m_{22}	c	d	f	g	h	0.4

						p
m_{23}	a	b	c	e	f	0.6
m_{24}	a	b	c	e	g	0.6
m_{25}	a	b	c	e	h	0.6
m_{26}	a	b	c	f	g	0.6
m_{27}	a	b	c	f	h	0.6
m_{28}	a	b	c	g	h	0.6
m_{29}	a	c	d	e	f	0.6
m_{30}	a	c	d	e	g	0.6
m_{31}	a	c	d	e	h	0.6
m_{32}	a	c	d	f	g	0.6
m_{33}	a	c	d	f	h	0.6
m_{34}	a	c	d	g	h	0.6
m_{35}	b	c	d	e	f	0.6
m_{36}	b	c	d	e	g	0.6
m_{37}	b	c	d	e	h	0.6
m_{38}	b	c	d	f	g	0.6
m_{39}	b	c	d	f	h	0.6
m_{40}	b	c	d	g	h	0.6
m_{41}	a	b	c	d	e	0.8
m_{42}	a	b	c	d	f	0.8
m_{43}	a	b	c	d	g	0.8
m_{44}	a	b	c	d	h	0.8

Donde el valor del error muestral es:

$$\sigma_{\bar{p}} = \sqrt{\frac{PQ}{n} \cdot \frac{N-n}{N-1}} = \sqrt{\frac{0.5 \cdot 0.5}{5} \cdot \frac{6-5}{6-1}} = 0.1$$

Por tanto, la distribución de las proporciones muestrales es la siguiente:

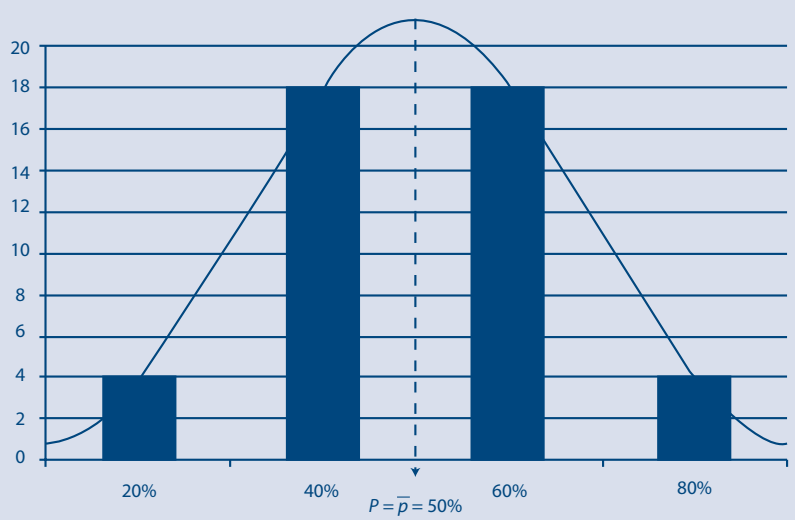


Figura 3.5

Estadísticos e-Books & Papers

Al igual que en la distribución de las medias muestrales, es posible determinar la probabilidad de que se pueda obtener una muestra con una proporción determinada, tal como se ejemplifica a continuación.

Problema resuelto

Con base en los estadísticos y parámetros del problema anterior, ¿cuál es la probabilidad de seleccionar una muestra al azar con un proporción de conductores multados entre 50% y 70%?

Solución

Considerando que:

$$P = 50\% \qquad \sigma_{\bar{p}} = 0.1 \qquad p = 70\%$$

entonces:

$$z = \left| \frac{0.7 - 0.5}{0.1} \right| = 2$$

Por lo que:

$$P(z = 2) = 0.4772 = 47.72\%$$

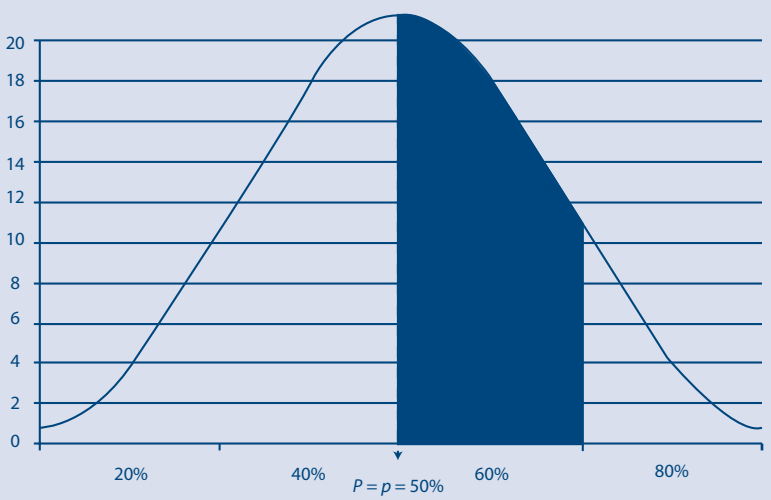


Figura 3.6

Alerta

Al igual que en la distribución de las medias muestrales, es posible determinar la probabilidad de que se pueda obtener una muestra con una proporción determinada.

3.6 Intervalos de confianza

Se puede señalar que es posible determinar los parámetros de la población siempre y cuando se conozcan todas las muestras representativas de la población; o sea que se estaría trabajando con la población misma. Sin embargo, en la práctica en muy pocas ocasiones es posible tener acceso a todas las muestras posibles de la población, por lo que las muestras presentarán ciertas limitaciones de diversos órdenes, por lo que se reconoce que existe error muestral.

Por lo anterior, no es posible conocer el valor del parámetro de manera sesgada; sin embargo, reconociendo que las medias muestrales se distribuyen normalmente, es posible determinar un rango dentro del cual se ubique el valor del mismo bajo ciertos niveles de probabilidad de ocurrencia, los cuales son mejor conocidos como niveles de confianza.

Los niveles de confianza se encuentran definidos en razón de la media y la desviación estándar.

En la siguiente gráfica se muestran los intervalos de confianza, así como las áreas bajo la curva (las probabilidades) que cada uno cubre.

Alerta

La distribución normal de las medias muestrales facilita determinar un rango dentro del cual se puede ubicar el valor de la media poblacional bajo niveles de probabilidad de ocurrencia, los cuales son mejor conocidos como niveles de confianza.

Estadísticos e-Books & Papers

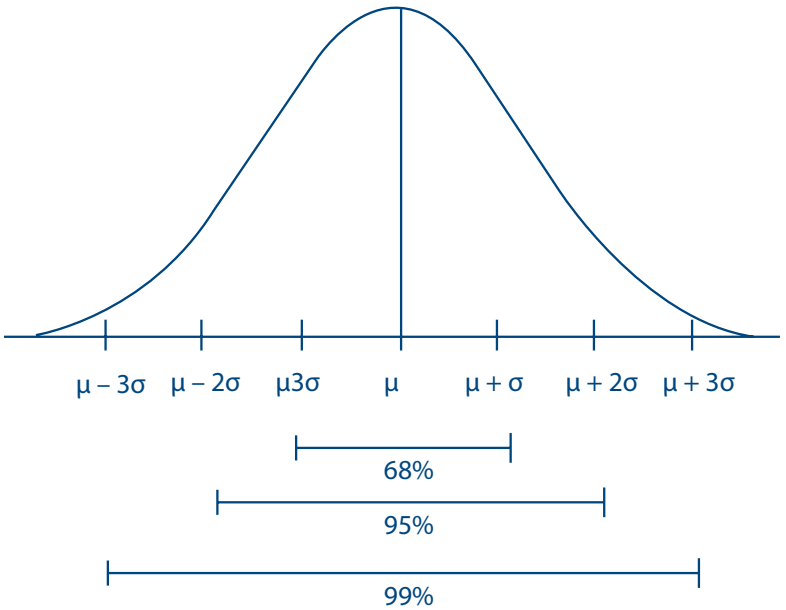


Figura 3.7

De manera práctica se puede señalar que bajo unidades estandarizadas se observa lo siguiente:

±1 Cuando $Z =$ El intervalo de confianza es de 68.27%

±2 Cuando $Z =$ El intervalo de confianza es de 95.45%

±3 Cuando $Z =$ El intervalo de confianza es de 99.73%

De manera complementaria los valores de Z para los diferentes niveles de confianza más comunes son los siguientes:

Nivel de confianza	*Z*
50.00%	0.6745
68.10%	1.0000
90.00%	1.6545
95.00%	1.9600
95.45%	2.0000
99.00%	2.5800
99.73%	3.0000

■ Estimación del intervalo de confianza para la media poblacional

Considerando el valor de una media muestral, así como el valor del error muestral, es posible determinar los valores de los límites del intervalo de confianza donde:

El valor del límite inferior está determinado como: $\overline{X} - z\,\sigma_{\overline{x}}$.
El valor del límite superior está determinado como: $\overline{X} + z\,\sigma_{\overline{x}}$.

Para ejemplificar, considérese el siguiente problema.

535
Estadísticos e-Books & Papers

Problema resuelto

Un estudio sobre la obesidad juvenil consideró tomar el peso a una muestra de 80 estudiantes, donde el peso promedio fue de 66 kilos, estimando una desviación estándar de la población estudiantil de 3 kilos.

Los responsables del estudio están interesados en conocer los límites de confianza de la media poblacional considerando niveles de confianza de 68.27%, 95.45% y 99.73%.

Solución

Los datos básicos para el cálculo son:

$$\overline{X} = 66 \qquad \sigma = 3 \qquad n = 80$$

Por lo que se procede a calcular el error muestral:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{80}} = 0.3354$$

Por tanto, los valores de los límites de los intervalos de confianza son los siguientes:

Nivel de confianza (NC)	z	Límite inferior $(\overline{X} - z\,\sigma_{\overline{x}})$	Límite superior $(\overline{X} + z\,\sigma_{\overline{x}})$
68.27%	1.0000	65.6646	66.3354
95.45%	2.0000	65.3292	66.6708
99.73%	3.0000	64.9938	67.0062

Sin embargo, cuando no es posible conocer o estimar el valor de la desviación estándar poblacional, los límites de confianza se pueden determinar calculando el error muestral considerando la desviación estándar muestral, por tanto:

$$S_{\overline{x}} = \frac{s}{\sqrt{n}}$$

Considerando que

$$\hat{S} = S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}}$$

Por tanto:

El valor del límite inferior está determinado como: $\overline{X} - z\ S_{\overline{x}}$.
El valor del límite superior está determinado como: $\overline{X} + z\ S_{\overline{x}}$.

Sin embargo, debe señalarse que $S_{\overline{x}}$ es buen estimador de $\sigma_{\overline{X}}$ siempre y cuando el valor de n sea grande, ya que el tamaño muestral permite que $\hat{S}$ se aproxime al valor de σ, pero n debe ser mayor que 30.

Alerta

Cuando no se conoce la desviación estándar poblacional es posible estimar los valores del intervalo de confianza utilizando el estimador de la desviación estándar muestral.

Estimación del intervalo de confianza para la proporción poblacional con base en la proporción muestral

El intervalo de confianza para la proporción de la población queda definido como:

El valor del límite inferior está determinado como: $\overline{P} - z\ \sigma_{\overline{p}}$.
El valor del límite superior está determinado como: $\overline{P} + z\ \sigma_{\overline{p}}$.
Recordando que el error estándar está definido como: $\sigma_{\overline{p}} = \sqrt{\frac{pq}{n}}$.

Estadísticos e-Books & Papers

Pero si de alguna forma no se puede estimar o determinar el valor del error estándar poblacional, se pueden determinar los valores de los límites de confianza del intervalo considerando el error estándar muestral:

$$S_p = \sqrt{\frac{pq}{n}}$$

Sin embargo, debe señalarse que S_p es un buen estimador de $\sigma_{\bar{p}}$ siempre y cuando el valor de n sea grande, pero mayor que 30.

Asimismo, debe acotarse que el valor máximo del producto (pq) se alcanza cuando ambos valen 50%.

Para ejemplificar lo anterior, considérese el siguiente problema.

Alerta

Cuando no se puede estimar o determinar el valor del error estándar poblacional, se pueden determinar los valores de los límites de confianza del intervalo considerando el error estándar muestral.

Problema resuelto

Los coordinadores del proceso de elección de los representantes estudiantiles de una facultad tienen registradas tres planillas: "Evolución", "Pro alumno" y "Mujeres en acción". Se sabe que de una muestra de 160 alumnos registrados en la facultad, 56 tienen la intención de votar por "Mujeres en acción".

¿Cuáles son los límites de confianza al 95% del porcentaje total de los alumnos de la facultad que votarán por la planilla en cuestión?

Solución

Se determina la proporción de la muestra:

$$p = \frac{56}{160} = 0.35 = 35\%$$

entonces $q = 1 - p = 1 - 0.35 = 0.65$

Por lo que, calculando el error estándar:

$$S_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.35)(0.65)}{160}} = 0.03771$$

Si el nivel de confianza es de 95% entonces $Z = \pm 1.96$, por lo que:

Nivel de confianza (NC)	z	Límite inferior $(p - zS_p)$	Límite superior $(p + zS_p)$
95.00%	1.9600	27.61%	42.39%

3.7 Determinación del tamaño muestral para estimar una media poblacional

Considerando que la distribución de las medias muestrales corresponde a una distribución normal, se puede señalar que el error muestral (E) refiere la diferencia entre una media muestral ($\overline{X}$) y la media poblacional (μ); por tanto, si se considera a partir de la media el error muestral define un intervalo de confianza:

Estadísticos e-Books & Papers

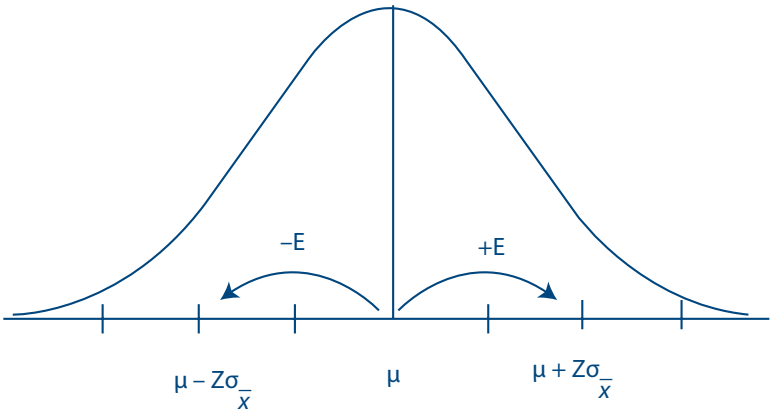


Figura 3.8

Por tanto:

$$e = \overline{X} - \mu = z\,\sigma_{\bar{x}}$$

Se obtiene que el intervalo de confianza queda definido como:

$$E = z\,\sigma_{\bar{x}} = z\frac{\sigma}{\sqrt{n}}$$

$$\mu \pm E = \mu \pm z\,\sigma_{\bar{x}} = \mu \pm z\frac{\sigma}{\sqrt{n}}$$

Por lo que, despejando *n*, se obtiene la fórmula para determinar el tamaño muestral:

$$E = z\frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{z \cdot \sigma}{E}$$

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2$$

En la fórmula anterior, si aplica, se puede estimar el valor de σ, así como el valor del error. Es importante indicar que el tamaño de la muestra también dependerá del nivel de confianza con el que se desee desarrollar la propuesta.

Considérese el siguiente problema.

Problema resuelto

Una tienda departamental considera realizar un estudio sobre el monto de ventas durante un periodo de crisis económica. Para ello desea determinar el tamaño de la muestra de las notas de ventas más conveniente si se estima una desviación estándar de $255, un margen de error de $50 y niveles de confianza de 95% y 99%.

Solución

$\sigma = \$255 \qquad E = \$50$

Si el nivel de confianza es de 95% entonces $Z = 1.96$, por tanto:

$$n = \left(\frac{z \cdot \sigma}{E}\right)^2 = \left(\frac{(1.96)\ (\$255)}{\$50}\right)^2 = 99.92 \approx 100 \text{ notas}$$

Alerta

El tamaño muestral para media población depende del nivel de confianza que se defina, siempre y cuando las demás variables permanezcan constantes.

Estadísticos e-Books & Papers

3.8 Determinación del tamaño muestral para estimar la proporción poblacional

De manera similar a lo expuesto en la sección anterior, también se puede determinar el tamaño de la muestra para proporciones partiendo de premisas similares como se expone a continuación. Se considera la siguiente gráfica.

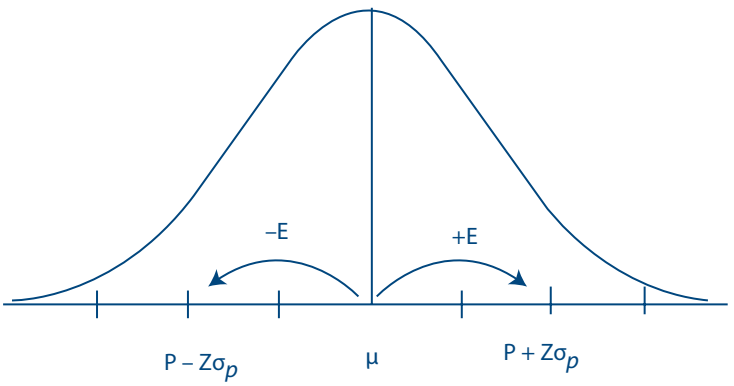


Figura 3.9

Por tanto:

$$E = p - P = z\sigma_p$$

Se obtiene que el intervalo de confianza queda definido como:

$$E = z\,\sigma_p = z\sqrt{\frac{P \cdot Q}{n}}$$

$$P \pm E = P \pm z\,\sigma_p = P \pm z\sqrt{\frac{P \cdot Q}{n}}$$

Por lo que despejando n se obtiene la fórmula para determinar el tamaño muestral:

$$E = z\sqrt{\frac{P \cdot Q}{n}}$$

$$E^2 = z^2\,\frac{P \cdot Q}{n}$$

$$n = \frac{z^2 \cdot P \cdot Q}{E^2}$$

Debiéndose considerar que el tamaño muestral dependerá en mucho del nivel de confianza propuesto, así como del producto $(P \cdot Q)$ debido, tal como se explicó anteriormente, a que el valor máximo del producto máximo se logra cuando $P = Q = 50\%$.

Para ejemplificar, considérese el siguiente problema.

Alerta

El tamaño muestral por proporciones depende del nivel de confianza y de los valores de *P* y *Q*.

Problema resuelto

El estudio de mercado para el lanzamiento de un nuevo modelo de tableta estima que 35% del público aficionado a la tecnología estaría dispuesto a comprarla. Para consolidar el estudio se desea estimar el tamaño muestral más conveniente para el estudio de preferencia de compra, por lo que se proponen niveles de confianza de 68%, 95% y 99%, así como un error de 5%.

Estadísticos e-Books & Papers

Solución

Si $P = 35\%$ entonces $Q = 1 - P = 65\%$, por tanto:

Nivel de confianza	Z	n (No. de personas)
68%	1.0000	$91.00 \approx 91.00$
95%	1.9600	$349.59 \approx 350.00$
99%	2.5800	$605.73 \approx 606.00$

Cabe señalar que entre más grande sea la muestra el error estándar se reduce.

3.9 Grados de libertad

Los grados de libertad pueden definirse como el número de elementos que pueden variar en su valor dentro de una colección.

Supóngase que se cuenta con tres variables: *K*, *L* y *M*. Si se dispone que el valor de *K* es 10, pero depende de los valores enteros que pueden tomar *L* y *M*, entonces los valores que pueden tomar son los siguientes:

$$K = 10 \text{ pero } K = L + M$$

A	10	9	8	7	6	5	4	3	2	9	0
B	0	1	2	3	4	5	6	7	8	1	10

O sea, de 3 variables pueden "variar 2"; entonces hay 2 grados de libertad.

Alerta

Los grados de libertad refieren el número de elementos que pueden variar en su valor dentro de una colección.

3.10 Intervalos de confianza para muestras pequeñas

Se considera como muestra pequeña aquella cuyo número de elementos es menor que 30.

Se ha determinado que la falta de elementos genera que el error muestral se incremente y por tanto la distribución de probabilidad no se ajuste a la normal. Por lo que, para compensar lo anterior, se utiliza la curva de distribución denominada *t* de Student, la cual es similar a la curva normal, pero un poco más robusta, y cuenta con las siguientes características:

- El valor del área bajo la curva es 1.
- Es simétrica con respecto a la media, pero es más extendida y chata en el centro en comparación con la curva de distribución normal.
- Al ser *t* una distribución más extendida, los valores de los límites de los intervalos de confianza son mayores en comparación con sus similares en la distribución normal.
- No hay una distribución única, sino un grupo de distribuciones, considerando que todas tienen como media 0 (cero), pero la desviación estándar dependerá del tamaño de la muestra.

La gráfica de la curva de distribución *t* es la siguiente:

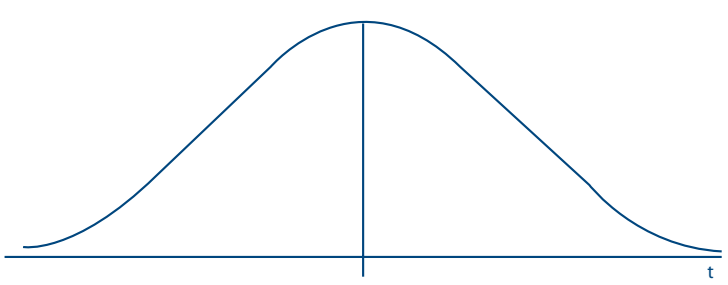


Figura 3.10

Estadísticos e-Books & Papers

Una de las principales premisas es que se desconoce el valor de la desviación estándar poblacional, por lo que se deben tener como referentes los estadísticos de la muestra. En consecuencia, al igual que en la curva normal, para determinar el valor de las unidades de *t* se utiliza la fórmula:

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

donde

$$\hat{S} = S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}}$$

Con base en el resultado obtenido de *t* se deberá considerar el nivel del intervalo de confianza.

Asimismo, para la determinación de los valores de los límites del intervalo de confianza para la estimación de la media poblacional con desviación estándar (σ) desconocida se aplica:

$$\overline{X} \pm t\frac{s}{\sqrt{n}}$$

Donde el valor de *t* depende del nivel de confianza propuesto y los grados de libertad, definidos como:

$$gl = n - 1$$

El nivel de confianza determina el área bajo la curva donde se puede ubicar el valor de la media poblacional, mientras que las áreas fuera de los límites del intervalo se denominan áreas de error (al error se le identifica con α). Para ubicar el valor de *t* se utilizan las tablas de distribución *t*, pudiendo ser de una cola (derecha o izquierda) o de dos colas. Por lo común se utiliza la tabla de distribución de dos colas, de manera que el error queda definido por $\frac{\alpha}{2}$.

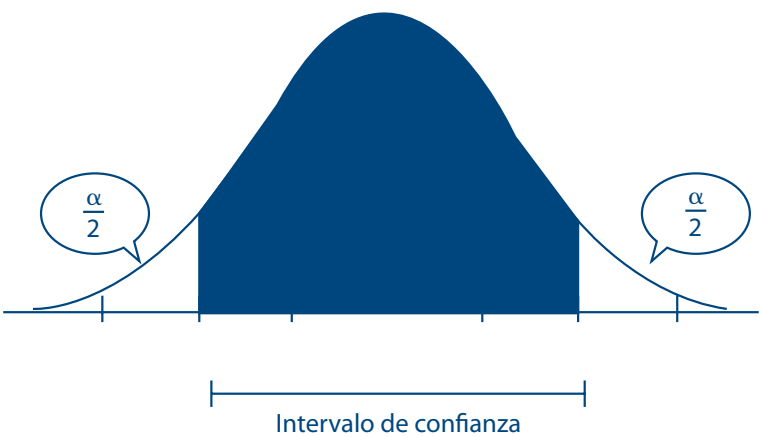


Figura 3.11

La definición del nivel de confianza se puede establecer de manera directa, por ejemplo 95%, pero también se puede definir el mismo tan solo señalando el valor del error, o sea que un valor de $\propto$ = 1% indica un intervalo de confianza de 99%.

Para ejemplificar lo anterior, revísese el siguiente problema.

Problema resuelto

Una obra de construcción cuenta con un laboratorio de campo para realizar pruebas para la medición de la resistencia de cilindros de concreto. De un concreto diseñado para 250 kg/cm^2 se cuenta con una muestra de 10 cilindros cuyas resistencias se muestran en la siguiente tabla.

541
Estadísticos e-Books & Papers

Problema resuelto *(continuación)*

Cil. #1	Cil. #2	Cil. #3	Cil. #4	Cil. #5
251	253	254	250	253
Cil. #6	**Cil. #7**	**Cil. #8**	**Cil. #9**	**Cil. #10**
252	250	253	251	253

Se desea estimar el valor de los límites de confianza de la media poblacional considerando un error de 5%.

Solución

Con base en los datos, el valor de la media y la desviación estándar muestrales son:

$$\overline{X} = 252.00$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}} = 1.4142$$

Grados de libertad:

$$gl = n - 1 = 10 - 1 = 9$$

Para $\alpha = 5\%$ (nivel de confianza de 95%) y $gl = 10$, el valor de t en dos colas es de 2.262, por lo que los valores de los límites del intervalo de confianza son:

Límite inferior: $\overline{X} - t\frac{s}{\sqrt{n}} = 252 - (2.262)\frac{1.4142}{\sqrt{10}} = 250.99$

Límite superior: $\overline{X} + t\frac{s}{\sqrt{n}} = 252 - (2.262)\frac{1.4142}{\sqrt{10}} = 253.01$

Alerta

Cuando el tamaño de una muestra es de 30 o más elementos, se considera que se puede distribuir con apego a la normal. Se considera como muestra pequeña aquella cuyo número de elementos es menor que 30 y su análisis debe apegarse a la distribución *t* de Student.

Estadísticos e-Books & Papers

Problemas para resolver

3.1 Con base en el conjunto:

$$U = \{18, 39, 66, 79, 90\}$$

determina las muestras de 3 elementos.

3.2 Con base en los resultados del problema anterior comprueba:

$$\overline{\overline{X}} = \mu \text{ y } \hat{\sigma}^2 = \hat{S}^2$$

3.3 Con base en el conjunto:

$$U = \{13, 21, 34, 45, 66\}$$

Determina las muestras consistentes y válidas de 4 elementos.

3.4 Con base en los resultados del problema anterior comprueba:

$$\overline{\overline{X}} = \mu \text{ y } \hat{\sigma}^2 = \hat{S}^2$$

3.5 Con base en el conjunto:

$$U = \{13, 24, 37, 56, 78, 98\}$$

determina muestras de 5 elementos.

3.6 Con base en los resultados del problema anterior comprueba:

$$\overline{\overline{X}} = \mu \text{ y } \hat{\sigma}^2 = \hat{S}^2$$

3.7 Considerando el planteamiento del problema 3.5, determina muestras de 4 elementos y comprueba.

3.8 Con base en los resultados del problema anterior comprueba.

$$\overline{\overline{X}} = \mu \text{ y } \hat{\sigma}^2 = \hat{S}^2$$

3.9 Considerando el conjunto:

$$P = \{200, 212, 234, 268, 294\}$$

determina las muestras, consistentes y válidas, de 3 elementos.

3.10 Con base en los resultados del problema anterior comprueba.

$$\overline{\overline{X}} = \mu \text{ y } \hat{\sigma}^2 = \hat{S}^2$$

3.11 Considerando el conjunto:

$$Q = \{10, 26, 84, 171, 192, 227, 258\}$$

determina las muestras, consistentes y válidas, de 5 elementos.

3.12 Con base en los resultados del problema anterior comprueba:

$$\overline{\overline{X}} = \mu \text{ y } \hat{\sigma}^2 = \hat{S}^2$$

3.13 Con base en los datos del problema 3.11 determina muestras de 3 elementos y comprueba que:

$$\overline{\overline{X}} = \mu \text{ y } \hat{\sigma}^2 = \hat{S}^2$$

SR **3.14** Una fábrica de focos para el hogar está diseñando un foco de 60 watts con componentes de bajo costo. Las pruebas iniciales de ajuste de la vida útil se desarrollaron con una población de prototipos de 6 focos cuyas horas de servicio fueron las siguientes:

Foco	a	b	c	d	e	f
Horas	19	34	53	89	97	116

El jefe de proyecto te asigna el desarrollo del análisis estadístico considerando muestras de 4 elementos a efecto de determinar la media y la desviación estándar poblacional, así como la comprobación de estas por estimadores insesgados.

3.15 Un despacho jurídico cuenta con 4 secretarias (2 por turno matutino o vespertino rolando turnos), las cuales llenan un mismo formato para el control de documentos del despacho, y que cometen el siguiente número de errores.

Secretaria	No. de errores
A	3
B	2
C	1
D	4

Si se seleccionan muestras de dos secretarias a partir de la población de 4 secretarias, determina y comprueba los parámetros poblacionales (media y desviación estándar) a través de los estimadores insesgados de las muestras.

SR **3.16** Una fábrica de rodamientos produce baleros de 2 pulgadas de diámetro. El departamento de control de calidad desea determinar el tamaño de la muestra de análisis al 90% de confianza, considerando una desviación estándar poblacional estimada de 0.04 pulgadas y un error de 0.06 pulgadas.

3.17 Una institución bancaria desarrolla un proceso de auditoría sobre los cortes de caja de sus diferentes sucursales. Si los responsables saben que la desviación estándar es de $120 y un error de $50, ¿cuál debe ser el tamaño de la muestra si considera un nivel de confianza de 99%?

SR **3.18** Un proceso de remodelación urbana afecta la hora de ingreso de los empleados de una empresa. El departamento de Recursos Humanos desea realizar un estudio para determinar si se incrementa de manera temporal el tiempo de tolerancia que es de 15 minutos. Se requiere determinar el tamaño de la muestra de los registros del tiempo de ingreso al 95% de confianza, considerando un error de 8 minutos.

SR **3.19** Una institución bancaria tiene por política que los ejecutivos de cuenta se conecten al sistema 5 minutos antes o después de la hora de apertura de las sucursales,

Estadísticos e-Books & Papers

que es a las 8:30 a.m. Los responsables de la evaluación de calidad en el servicio (SERVQUAL) desean determinar el tamaño de la muestra al 95% y 99% de confianza de los registros de conexión al sistema, considerando que la desviación estándar es de 2 minutos 40 segundos y un error de 1 minuto 15 segundos.

3.20 Una aerolínea regional desea realizar un estudio sobre el número de pasajeros en lista de espera por mes. Se sabe que el número de pasajeros en la lista fluctúa en más o menos 5 por vuelo. Se desea estructurar una muestra sobre las listas de espera al 99% de confianza, considerando un margen de 2 pasajeros.

3.21 Un estudio sobre pacientes diabéticos demuestra que requieren aplicarse después de los alimentos más o menos 4 unidades de una nueva insulina. Si se considera un error de 1.5 unidades, se desea determinar el tamaño de la muestra de pacientes para desarrollar el estudio sobre el nuevo fármaco a un nivel de confianza de 90%.

3.22 Con base en los datos del problema anterior, se considera oportuno desarrollar una investigación con una muestra al 99% de confianza.

3.23 La evaluación sobre el índice de reprobación de una materia universitaria indica que 45% de los alumnos reprueba. Se requiere conformar una muestra de los alumnos considerando 5% de error y un nivel de confianza de 95.45% con el propósito de desarrollar las medidas que permitan reducir el índice de reprobación.

3.24 Una importante cadena de supermercados sabe que en 25% de las operaciones los cajeros no cuentan con el dinero fraccionario para dar el cambio a los clientes. Los responsables de la tesorería de la cadena desean establecer un estudio inferencial sobre esta situación, estructurando una muestra al 99% de confianza y considerando un error de 10%.

3.25 Una cadena de tlapalerías sabe que en 78% de las ventas de thinner no se ofrece el litro exacto, o sea que se ofrece más o menos producto. Se desea desarrollar un proceso de análisis sobre los inventarios del solvente, considerando muestras al 90%, 95%, 95.45%, 99% y 99.73% y con un nivel de error de 10%.

3.26 Una fábrica de acumuladores para automóviles encuentra que la vida media de una muestra de 26 acumuladores es de 39 meses con una desviación estándar de 5 meses. Si se consideran grados de confianza de 90%, 95% y 99%, ¿cuáles son los valores de los límites de confianza donde se puede encontrar la media poblacional?

3.27 Con base en los datos de la muestra, procede a ordenar y determinar los límites de confianza para la media poblacional al 95% y 99%.

142	161	150	145	130	184	176	154	149	178
112	90	162	77	198	181	116	100	165	199

3.28 Con base en los datos del problema anterior, determina los límites de los intervalos de confianza al 50% y 68%.

3.29 Con base en los datos de la muestra, procede a ordenar y determinar los límites de confianza para la media poblacional al 50% y 68%.

1 234	1 178	1 340	1 189	1 567
1 645	1 934	1 937	1 756	1 111
1 023	1 404	1 023	987	1 032

3.30 Con base en los datos del problema anterior, determina los límites de los intervalos de confianza al 95.45% y 99%.

3.31 Considerando datos de la muestra, procede a determinar los límites de confianza al 50%, 68% y 90%.

169	128	140	176	161	119	164	154
140	136	154	136	148	144	175	148
136	162	130	141	123	159	151	150
146	139	160	146	136	148	144	152
145	142	156	145	128	150	135	161

3.32 Con base en los datos del problema anterior, determina los límites de los intervalos de confianza al 95%, 95.45% y 99%.

3.33 A partir de la colección de datos, determina el valor de los límites de los intervalos de confianza al 95%, 95.45% y 99%.

169	128	140	176	161	119	164	154
140	136	154	136	148	144	175	148
136	162	130	141	123	159	151	150
146	139	160	146	136	148	144	152
145	142	156	145	128	150	135	161

3.34 Considerando datos de la muestra, procede a determinar los límites de confianza al 50%, 68% y 90%.

3.35 Una empresa de elementos prefabricados de madera procedió a clasificar los excedentes de bastidores de una pulgada de espesor por su longitud en centímetros. Donde el detalle del inventario se expone a continuación.

25	37	47	60	74
34	38	52	63	66
27	38	49	61	72
46	64	70	44	45
41	53	62	67	42
45	59	71	72	60
50	58	51	56	52
45	49	52	53	57

Considerando esta información, procede a determinar el intervalo de confianza de la media poblacional, considerando el tamaño muestral en cuanto a la aplicación de la distribución normal o *t* de Student bajo niveles de 68% y 95.45%.

Estadísticos e-Books & Papers

3.36 Al término de un proyecto de construcción, una empresa de proyectos electromecánicos procede a realizar el levantamiento del inventario de los tramos de cable de 500 MCM considerando su largo en pulgadas. Los responsables del departamento de Ingeniería de Procura (Compras) estructuran la muestra que se detalla a continuación.

37	38	38	64	53	59	58	49
74	66	72	45	42	60	52	57
60	63	61	44	67	72	56	53
25	34	27	46	41	45	50	45
65	73	39	70	54	73	71	61
47	52	49	70	62	71	51	52

Considerando esta información, procede a determinar el intervalo de confianza de la media poblacional, considerando el tamaño muestral en cuanto a la aplicación de la distribución normal o *t* de Student bajo niveles de error de 10% y 5%.

3.37 Considerando el planteamiento del problema anterior, determina el valor de los límites del intervalo de confianza al 99%.

3.38 Un fabricante de focos desarrolla un estudio sobre la vida útil de un foco ahorrador de 20 watts a través de una muestra de 15 unidades, donde las horas de servicio de cada foco son las siguientes:

520	521	511	513	510
513	522	500	521	495
496	488	500	502	512

Por lo que, en atención a la información mostrada, determina los intervalos de confianza sobre la media poblacional al 90%, 95% y 99%.

3.39 Determina los valores de los intervalos de confianza para la media poblacional al 50% y 68% a través de los estadísticos de la muestra que se expone a continuación:

11	27	34	42
22	56	35	56
45	67	37	67
78	73	66	98
65	90	78	95

3.40 Considerando los datos del problema anterior, determina los valores de los límites de los intervalos de confianza cuando $\propto = 5\%$ y $\propto = 1\%$.

3.41 Una empresa internacional de control de calidad en materiales de construcción obtiene una muestra de 10 cilindros de concreto, los cuales son ensayados a compresión obteniendo las siguientes resistencias en kg/cm^2.

Cil. 1	Cil. 2	Cil. 3	Cil. 4	Cil. 5	Cil. 6	Cil. 7	Cil. 8	Cil. 9	Cil. 10
248	252	254	251	249	250	252	255	253	248

Determina el intervalo de confianza de la media poblacional al 90%, 95% y 99%.

3.42 El gerente de una tienda de refacciones para tornos de control numérico desea estimar la cantidad media de venta de refacciones. Una muestra de 20 facturas (en dólares) muestra las siguientes cantidades:

$48.16	$52.64	$51.35	$23.78	$61.83
37.92	61.46	51.45	46.94	43.88
49.17	46.82	50.82	58.84	54.86
42.22	48.59	52.68	41.86	61.69

Si eres nombrado asesor de operaciones del gerente, procede a determinar lo siguiente:

¿Cuál es el mejor intervalo de confianza para la media poblacional, considerando niveles de error de 10%, 5% y 1% en prueba *t* de dos colas?

3.43 Una fábrica de grasa grafitada para maquinaria envasa la grasa en latas de 500 g. El área de control de calidad realiza una revisión sobre el proceso de envasado, para ello consideró una muestra de 10 paquetes con los siguientes pesos en gramos:

497	492	510	505	494
503	500	492	501	498

A las gerencias de producción y de control de calidad les interesa determinar el estimado de la media poblacional al 90%, 95% y 99%.

3.44 Una fábrica de bastidores de madera para muebles mide la resistencia a la compresión de 2 muestras de bastidores de $\frac{1}{3}$ pulg de espesor, donde las pruebas de laboratorio arrojaron los siguientes resultados en kg/cm^2.

Muestra *A*		Muestra *B*
10.14		10.56
11.24		10.66
10.38		10.45
10.65		10.90
10.80		11.00

Como responsable del análisis de producción, procede a desarrollar los cálculos necesarios de manera que se determinen los intervalos de confianza al 90% para la media poblacional de cada muestra.

3.45 Con base en el planteamiento del problema anterior, determina el intervalo de confianza al 95% para la media poblacional considerando la muestra completa.

3.46 Con base en el planteamiento del problema 3.36, procede a desarrollar los cálculos necesarios de manera que se determinen los intervalos de confianza al 99% para la media poblacional de cada muestra.

3.47 Con base en el planteamiento del problema anterior, determina el intervalo de confianza al 90 y 95% para la media poblacional considerando la muestra completa.

Estadísticos e-Books & Papers

3.48 Con base en el planteamiento del problema de la fábrica de focos para el hogar, determina el intervalo de confianza para la media poblacional al 95%.

Foco	a	b	c	d	e	f
Horas	19	34	53	89	97	116

3.49 Un análisis inferencial considera una media muestral de 8, una desviación estándar poblacional de 2, un tamaño muestral de 44, y límites de confianza de LI = 7.41 y LS = 8.59. Considerando una distribución normal, determina:

a) El nivel de confianza del análisis.

b) El valor del error muestral.

PROBLEMAS RETO

1

El Ministerio de Finanzas señala que el llenado de un formato vía Internet por parte de los ciudadanos promedia 120 minutos. Sin embargo, una muestra de 40 personas expuso que el tiempo promedio es de 80 minutos. Bajo condiciones de aproximación a la normal, al director de la oficina le interesaría conocer lo siguiente:

a) ¿Cuál es el valor del error entre las medias?

b) ¿Cuál es el valor de la desviación estándar poblacional al 68%, 95% y 99% de confianza?

2

Con base en el planteamiento del problema anterior, al director le interesa responder de manera adicional las siguientes interrogantes:

a) Al 99% de confianza, ¿cuál es la probabilidad de que los contribuyentes tarden más de 120 minutos?

b) Al 99% de confianza, ¿cuál es la probabilidad de que los contribuyentes tarden entre 120 y 150 minutos?

c) Al 99% de confianza, ¿cuál es la probabilidad de que tarden más de 150 minutos?

REFERENCIAS

Berenson, Mark L. y David M. Levine (1996). *Estadística básica en administración* (6a. ed.). México: Prentice Hall Hispanoamericana.

Hines, William W., Douglas C. Montgomery, David M. Goldsman y Connie M. Borror (2009). *Probabilidad y estadística para ingenieros* (4a. ed.). México: Patria, 3a. reimpresión.

Kohler, Heinz (1996). *Estadística para negocios y economía* (1a. ed.). México: Compañía Editorial Continental.

Lind, Douglas, William G. Marchal y Samuel A. Wathen (2005). *Estadística aplicada a los negocios y la economía* (12a. ed.). México: McGraw-Hill.

Mendenhall, William (1999). *Estadística para administradores* (2a. ed.). México: Grupo Editorial Iberoamérica.

Quevedo, Urias Héctor y Blanca Rosa Pérez Salvador (2008). *Estadística para Ingeniería y Ciencias*. México: Grupo Editorial Patria.

Estadísticos e-Books & Papers

DIRECCIONES ELECTRÓNICAS

Portal académico CCH-Distribuciones de Probabilidad
[http://portalacademico.cch.unam.mx/alumno/sitiosdeinteres/matematicas/estadistica2]
Distribución normal
[http://es.wikipedia.org/wiki/Distribuci%C3%B3n_normal]
t de Student
[http://jacroman.blogspot.mx/2009/12/distribucion-t-student.html]
Tabla *t* de Student
[ftp://ftp.fcien.edu.uy/Programas/ftp/bioe2007/archivos/tablat.pdf]

547
Estadísticos e-Books & Papers

UNIDAD 4

Análisis estadístico de experimentos

OBJETIVOS

- Entender la importancia de la aplicación de la estadística inferencial al análisis de los resultados derivados de un experimento.
- Comprender y distinguir los conceptos de experimento y prueba.
- Comprender el concepto de tratamiento.
- Distinguir la diferencia entre variable dependiente y variable independiente.
- Comprender el concepto de hipótesis estadística.
- Distinguir la diferencia entre hipótesis nula e hipótesis alternativa.
- Comprender la importancia de la aplicación de la distribución *t* de Student para el análisis de pruebas de hipótesis estadísticas.
- Entender el concepto de análisis de la varianza (ANOVA).
- Comprender la aplicación de la distribución de la *F* de Fisher para el análisis de las pruebas de hipótesis estadísticas.
- Entender el concepto de discriminantes para el análisis de diferenciación de resultados.

¿QUÉ SABES?

- ¿Cuál es la trascendencia de la aplicación de la estadística inferencial al análisis de los resultados derivados de un experimento?
- ¿Cuál es la diferencia entre un experimento y una prueba?
- ¿Qué es un tratamiento?
- ¿Cuál es la diferencia entre una variable dependiente y una independiente?
- ¿Qué es una hipótesis estadística?
- ¿Cuál es la diferencia entre una hipótesis nula y una hipótesis alternativa?
- ¿Cuál es la importancia de la aplicación de la distribución *t* de Student para el análisis de pruebas de hipótesis estadísticas?
- ¿Qué es análisis de la varianza (ANOVA)?
- ¿En qué forma la distribución de la *F* de Fisher apoya el análisis de las pruebas de hipótesis estadísticas?
- ¿Qué son los discriminantes para el análisis de diferenciación de resultados de un experimento?

Estadísticos e-Books & Papers

4.1 Introducción

Alerta

Los elementos sujetos a observación y medición en un experimento son las variables.

En el transcurso de las investigaciones científicas e industriales se realizan procesos planificados de investigación y desarrollo en los cuales se llevan a cabo experimentos. Los resultados obtenidos en los procesos de experimentación deben ser organizados, computados y representados mediante gráficas para proceder a su interpretación y análisis mediante la estadística inferencial con el fin de comprobar, refutar o generar conocimientos.

4.2 Concepto de experimento

Alerta

Las variables en un experimento, de manera básica, se clasifican en dependientes e independientes.

Experimento es el conjunto de actividades interrelacionadas, las cuales se realizan con el propósito de practicar o replicar un fenómeno de manera planificada y controlada para obtener un resultado esperado en razón al número y magnitud de las variables que lo estructuran.

De manera básica se puede señalar que las variables son los elementos sujetos a observación y medición durante un experimento, si consideramos que una o más variables son manipuladas por los investigadores para generar un resultado específico. Por consecuencia, a las variables se les puede clasificar en las siguientes.

a) **Variables independientes.** Son aquellas manipuladas a discreción por los investigadores para generar un resultado específico y se les representa con la letra *x*.

b) **Variables dependientes.** Son las que muestran el resultado o fenómeno en estudio y están representadas por la letra *y*.

Alerta

La estructura básica de las hipótesis de investigación científica es la hipótesis estocástica: "*si x* entonces *y*".

Lo anterior dispone que las variables independientes causan o explican a las dependientes. Este argumento da origen a la estructura básica de las hipótesis de investigación científica, o sea las hipótesis estocásticas: "si *x* entonces *y*".

En resumen, durante una investigación científica de orden aplicado a la industria, los investigadores manipularán las variables *x* buscando generar un resultado esperado referido por la variable *y*.

Ante lo expuesto, se puede señalar que en muy contadas ocasiones los resultados esperados de un experimento se obtienen en primera instancia. Ante ello los investigadores deben realizar ajustes durante el experimento, así como repetirlos bajo diferentes condiciones y estructura de las variables a efecto de observar el comportamiento del fenómeno en estudio. A este concepto se le conoce como tratamiento, es decir, los investigadores pueden repetir un experimento a partir de diferentes tratamientos.

Alerta

Una hipótesis estadística expone una conjetura o condición sobre un parámetro de la población, la cual se comprueba o refuta de acuerdo con el análisis de los estadísticos de la muestra.

Asimismo, cuando se realiza (ejecuta) un tratamiento se le denomina ensayo o prueba, por lo que un experimento puede constar de varios tratamientos y ensayos, de manera que los resultados de un tratamiento pueden ser contrastados con los de otro u otros tratamientos a efecto de obtener conclusiones referentes al fenómeno en estudio, incluso comparar uno o varios tratamientos contra otro considerado como testigo. A lo expuesto se le conoce como estudios comparativos.

4.3 Hipótesis estadísticas

Son argumentos que exponen una condición sobre un parámetro de la población, el cual está sujeto a verificación mediante la evidencia cuantitativa derivada del análisis estadístico sobre una o varias muestras.

Las hipótesis estadísticas de manera general quedan definidas como sigue:

- **Hipótesis nula (H_o).** Es una afirmación sobre un parámetro poblacional, el cual se trata de anular mediante la evidencia.
- **Hipótesis alternativa (H_1).** Se refiere a la negación del argumento expuesto en la hipótesis nula, la cual se valida en el momento que se refuta H_o.

Alerta

El análisis de las hipótesis estadísticas puede generar que se cometan errores tipo 1 o tipo 2.

Sin embargo, puede llegar a suceder que existan condiciones que impidan el buen desarrollo de un experimento, así como el proceso de análisis estadístico que dé origen a los llamados errores estadísticos, los cuales pueden ser de dos tipos.

- **Tipo 1 o α.** Se rechaza H_o cuando es verdadera.
- **Tipo 2 o β.** Se acepta H_o cuando es falsa.

Estadísticos e-Books & Papers

4.4 Aplicación de la distribución *t* de Student en el análisis de experimentos

Debe señalarse que la realización de experimentos tanto académicos como industriales depende en gran medida de la disponibilidad de recursos y el tiempo para llevarlos a cabo. Puede ocurrir que en los experimentos se utilicen muestras pequeñas o prototipos, es decir, que se cuente con un número limitado de elementos.

Lo anterior expone que se manejen tamaños muestrales pequeños, por lo que el análisis de experimentos se fundamenta en la distribución *t* de Student.

4.5 Estudios comparativos simples

La profundidad en cuanto al análisis estadístico de un experimento depende en gran medida de su propósito, pero en esencia se procura realizar varios tratamientos cuyos resultados se comparen a efecto de obtener conclusiones.

Figura 4.1

De manera general, los estudios comparativos simples exponen las siguientes hipótesis estadísticas:

- $H_o = \mu_1 = \mu_2 = \ldots = \mu_n$ No hay diferencia entre las medias.
- $H_1 = \mu_1 \neq \mu_2 \neq \ldots \neq \mu_n$ Hay diferencia entre las medias.

El planteamiento de las hipótesis es simple de explicar, ya que en cada muestra o grupo se realiza un tratamiento distinto, de manera que si las medias no difieren en su valor, es posible que los tratamientos no generen el efecto o resultado esperado y por consecuencia se valida la hipótesis nula.

De manera práctica, la prueba de hipótesis consiste en contrastar los resultados de los tratamientos por pares de muestras. De modo que a partir de una varianza combinada,

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Donde

S_1^2 = varianza de la muestra uno
S_2^2 = varianza de la muestra dos
n_1 = tamaño de la muestra uno
n_2 = tamaño de la muestra dos

Se determina el valor de *t* calculada identificada por t_0,

$$t_0 = \frac{\overline{Y}_2 - \overline{Y}_1}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Alerta

En la teoría del análisis de experimentos el término grupo se utiliza como sinónimo de muestra.

Estadísticos e-Books & Papers

Donde el valor de t_0 se deberá contrastar con una t_{tab} teórica o tabular definida como: $t_{tab}\left(\frac{\alpha}{2}, gl\right)$, donde los grados de libertad se definen como: $gl = (n_1 + n_2) - 2$. De manera gráfica la t_{tab} tabular define los límites de las áreas entre H_o y H_1.

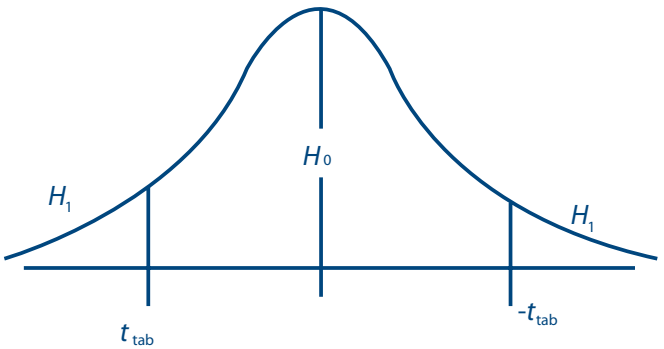


Figura 4.2

De manera que si,

$$t_0 < t_{tab} \text{ o } t_0 > t_{tab} \text{ se acepta } H_1.$$

Para ejemplificar lo anterior considérese el siguiente problema.

Problema resuelto

Una empresa que manufactura armas crea un nuevo diseño de cartuchos para escopeta. Con el fin de darle mayor estabilidad e impermeabilidad al cuerpo del cartucho, experimenta con dos tipos de resinas plásticas cuyo grosor origina que el peso del cartucho varíe.

Realiza un análisis comparativo simple para determinar si existe diferencia en los dos tipos de cartuchos, dependiendo del tipo de resina del cuerpo si se considera el peso, en gramos, de seis cartuchos por muestra.

Resina A (RA)	Resina A-1 (RA1)
32.30	31.89
31.90	32.10
31.45	32.34
32.80	32.40
33.01	32.30
32.78	31.99

Solución

Considérese que x = Resina, y = Peso y que:

$$H_o = \mu_1 = \mu_2$$
$$H_1 = \mu_1 \neq \mu_2$$

551
Estadísticos e-Books & Papers

Solución *(continuación)*

Por lo que, si seguimos el procedimiento de cálculo, se determinan los valores de la media y la varianza de cada muestra.

$$\overline{Y}_1 = 32.37 \qquad \overline{Y}_2 = 32.18$$

$$S^2 = \frac{\Sigma(Yi-\overline{Y})^2}{n-1}$$

$$S_1^2 = 0.367 \qquad S_2^2 = 0.043$$

El valor de la varianza combinada es:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(6-1)(0.367)+(6-1)(0.043)}{6+6-2} = 0.2025$$

Por lo que el valor de t_0 es:

$$t_0 = \frac{\overline{Y}_2 - \overline{Y}_1}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{32.37-32.17}{0.45274\sqrt{\frac{1}{6}+\frac{1}{6}}} = 0.7651$$

Ya que el valor de t_{tab}:

$$\alpha = 5\% \qquad \frac{\alpha}{2} = 0.025$$

$$gl = (n_1+n_2)-2 = 12-2 = 10$$

$$t_{tab}\left(\frac{0.05}{2},\ 10\right) = 2.228$$

Al contrastar los valores de manera gráfica:

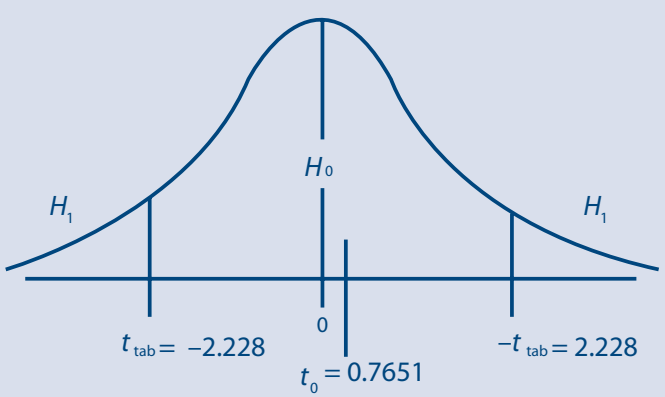


Figura 4.3

Los cálculos exponen que el valor de t_0 se ubica en el área correspondiente a la hipótesis nula. Es decir, las resinas no generan diferencias entre los pesos de los cartuchos.

Estadísticos e-Books & Papers

4.6 Estudios comparativos basados en pruebas de hipótesis sustentados en el análisis de la varianza (ANOVA) de un factor con una muestra por grupo

Este tipo de estudios se fundamenta en el análisis de las posibles fuentes de variación en los resultados de un experimento con respecto a la media de la variable dependiente, como son las siguientes.

a) El error aleatorio con origen en el proceso de medición.

b) Errores en los factores controlados, por ejemplo:

- Desarrollo del método propuesto para la experiencia.
- Fallas en los equipos e instrumentos.
- Condiciones del medio ambiente.
- Condiciones del analista.

El análisis de ANOVA permite realizar estudios comparativos entre más de dos muestras (grupos) que han sido sometidas a diferentes tratamientos con base en establecer qué tan grande es la variación de los datos de una muestra con respecto a su media (variación dentro de grupos), con respecto a la variación entre las medias muestrales y con respecto a la media general (variación entre grupos).

La premisa anterior permite definir las siguientes hipótesis estadísticas:

- $H_o = \mu_1 = \mu_2 = ... = \mu_n$ No hay diferencia entre las medias.
- $H_1 = \mu_1 \neq \mu_2 \neq ... \neq \mu_n$ Hay diferencia entre las medias.

La prueba de hipótesis bajo ANOVA se fundamenta en la distribución *F* de Fisher, la cual es una distribución de probabilidad continua, también conocida como distribución *F* de Snedecor (por George Snedecor).

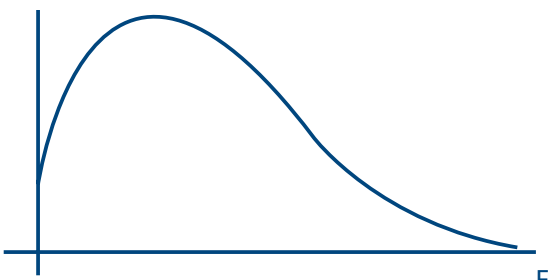


Figura 4.4

La distribución *F* de Fisher es la más conocida por su aplicación a las pruebas de hipótesis estadísticas y el problema más simple del análisis de varianza, y se utiliza debido a que la hipótesis propone que las medias de múltiples poblaciones son iguales si están normalmente distribuidas y con la misma desviación estándar.

De manera concreta, la prueba de hipótesis propuesta consiste en contrastar el valor de una *F* calculada (*Fc*) con respecto al valor de una *F* tabular o teórica (*Ft*); de hecho, tal como se aprecia en la gráfica siguiente, cuando $Ft < Fc$ se cumple H_1.

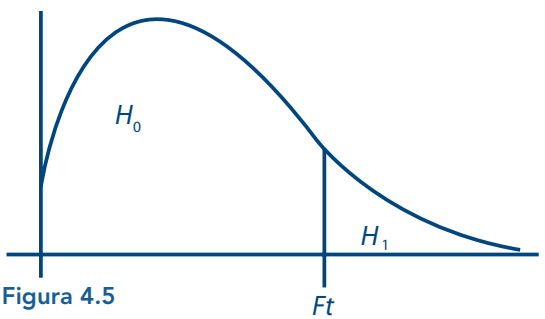


Figura 4.5

Alerta

El término ANOVA es el acrónimo de *Analysis Over Variance*.

Alerta

La prueba de hipótesis por medio de ANOVA se fundamenta en la distribución *F* de Fisher.

Estadísticos e-Books & Papers

El procedimiento de cálculo de ANOVA parte de la determinación de valor de las variaciones. El valor de la variación dentro de grupos se calcula así:

$$Var_{DG} = \Sigma_{i=1}^{n}(X_i - \overline{X})^2$$

Mientras que el valor de la variación entre grupos se determina por la fórmula:

$$Var_{EG} = \Sigma_{i=1\,n}^{n}(X_n - \overline{\overline{X}})^2$$

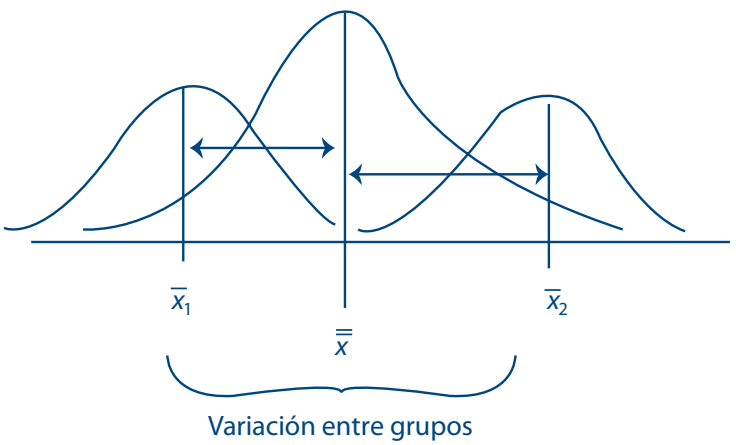


Figura 4.6

En consecuencia, el valor de la *F* calculada se determina así:

$$Fc = \frac{\dfrac{Var_{EG}}{N-1}}{\dfrac{Var_{DG}}{N-Nm}}$$

Donde

N = el número total de datos aportados por las muestras.
Nm = el número de muestras.
Var_{EG} = la variación entre grupos.
Var_{DG} = la variación dentro de grupos.

Para determinar el valor de la *F* tabular se requiere conocer el valor del error (α), así como los grados de libertad a efecto de ubicar el valor en la tabla correspondiente:

$$Ft\,[\alpha, (Nm - 1), (N - Nm)]$$

donde los valores de los grados de libertad están definidos así:

Nm – 1 = denominados grados de libertad del numerador.
N – *Nm* = denominados grados de libertad del denominador.

Para ejemplificar lo anterior se presenta el siguiente problema.

554
Estadísticos e-Books & Papers

Problema resuelto

Realiza una prueba de hipótesis por ANOVA del experimento de las resinas para cartuchos de escopeta a efecto de determinar si existe diferencia en el peso de los cartuchos con base en el tipo de resina utilizada. Considera un nivel de confianza del 95%.

Solución

Se procede a calcular las variaciones dentro de grupos y entre grupos.

$\overline{X}_{RA} = 32.373$ $\overline{X}_{RA1} = 32.170$

RA	RA1	$\left(X^i - \overline{X}_{RA}\right)^2$	$\left(X^i - \overline{X}_{RA1}\right)^2$
32.30	31.89	0.005329	0.0784
31.90	32.10	0.223729	0.0049
31.45	32.34	0.851929	0.0289
32.80	32.40	0.182329	0.0529
33.01	32.30	0.405769	0.0169
32.78	31.99	0.165649	0.0324
	Σ	1.834734	0.2144

De manera que

$Var_{DG} = 1.8313 + 0.2144 = 2.0457$

$Var_{EG} = 6(32.373 - 32.2716)^2 + 6(32.17 - 32.2716)^2 = 0.1236$

$$Fc = \frac{\frac{Var_{EG}}{Nm-1}}{\frac{Var_{DG}}{N-Nm}} = \frac{\frac{0.1236}{2-1}}{\frac{2.0457}{12-2}} = 0.6037$$

De la tabla de la distribución *F* de Fisher para $\alpha = 5\%$:

$Ft\,[0.5, 1,10] = 4.9650$

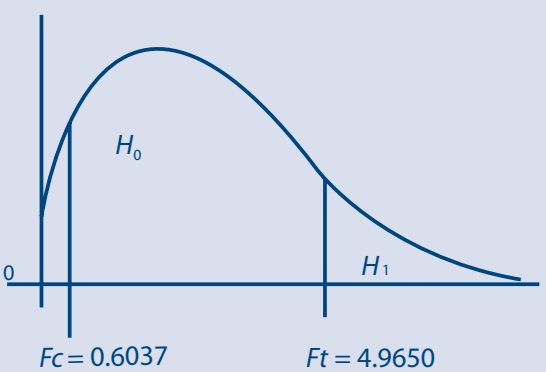


Figura 4.7

Como $Ft > Fc$ entonces se comprueba H_o, lo que significa que no hay diferencia entre las medias, o sea que el tipo de resina del cuerpo de los cartuchos no provoca que haya variaciones en sus pesos.

555
Estadísticos e-Books & Papers

Limitaciones de las pruebas de hipótesis fundamentadas en ANOVA

El resultado de la prueba tan solo señala si existe o no diferencia, pero en caso de que se cuente con más de dos muestras y la prueba señale que hay diferencia, la prueba no discrimina cuál muestra es, por lo que es necesario llevar a cabo otro tipo de procedimientos de cálculo para discriminar entre las muestras.

4.7 Estudios comparativos basados en ANOVA de dos factores

Este tipo de estudios sirven para evaluar el comportamiento individual y el conjunto de dos o más variables (factores) sobre una variable dependiente cuantitativa. El análisis permite observar la incidencia o efectos de cada variable en lo individual, así como la interacción entre ambas; es decir, se estudian tres efectos. Por señalar si se contará con el análisis de tres variables (factores), entonces los efectos serían siete: tres individuales, tres por pares de factores y uno de interacción entre ellos.

Para precisar considérese el caso que se cuente con *m* muestras con tamaño *n* dispuestas de forma matricial, donde uno de los factores en estudio, denotado por la letra *a* estará referido por los renglones, mientras que el segundo factor, denominado *b*, estará referido por las columnas.

Alerta

Al ANOVA de dos factores también se le conoce como prueba ANOVA de dos vías.

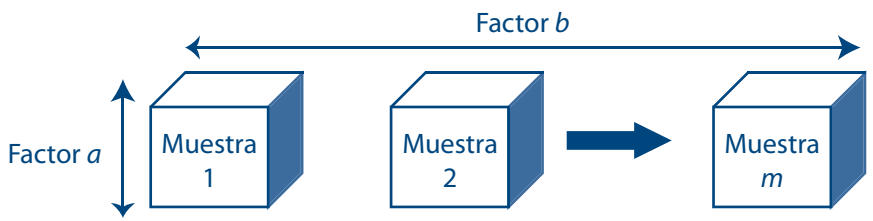


Figura 4.8

De manera práctica el procedimiento de cálculo se expone mediante el siguiente ejemplo.

Una empresa pretende innovar en el proceso de fabricación de uno de sus productos, por lo que prueba tres presentaciones diferentes del insumo principal: barra, granulado y en polvo. Ensaya cada presentación en las ocho máquinas que tiene. La gerencia de manufactura se interroga si existen diferencias significativas en el nivel de producción dependiendo del tipo de presentación del insumo y del tipo de máquina.

Tabla 4.1

Factor *b* Tipo de insumo

Factor *a* Máquina	Barra	Granulado	En polvo
Máquina 1	28	31	30
Máquina 2	26	29	31
Máquina 3	26	27	29
Máquina 4	24	27	28
Máquina 5	20	25	27
Máquina 6	22	23	28
Máquina 7	24	25	27
Máquina 8	25	26	29

Con el enfoque de ANOVA de dos factores el tamaño muestral deberá entenderse por el número de elementos que se ubican bajo la relación renglón-columna, que en el caso del ejemplo es uno, por lo que el tamaño total de observaciones está definido como: $N = abn$.

Estadísticos e-Books & Papers

Por lo que para calcular el ANOVA se deben obtener las sumas de los cuadrados de los efectos de cada factor.

a) Para el factor *a* (renglones).

$$SC_a = \frac{\sum_{j=1}^{b} Sum\_Col_j^2}{b \cdot n} - \frac{Sum\_Total^2}{N}$$

b) Para el factor *b* (columnas).

$$SC_b = \frac{\sum_{i=1}^{b} Sum\_Reng_i^2}{a \cdot n} - \frac{Sum\_Total^2}{N}$$

c) Se calcula el valor de los cuadrados totales.

$$SC_{Tot} = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij}^2 - \frac{Sum\_Total^2}{N}$$

d) Se calcula el valor de los cuadrados del error o simplemente el error.

$$SC_{Error} = SC_{Tot} - [SC_a + SC_b]$$

Con base en los parámetros anteriores se estructura el cuadro de análisis.

Tabla 4.2 Análisis de varianza

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F_{calc}	F_{tab}	Prob.
Factor *a*	SC_a	$(a-1)$	$\frac{SC_a}{a-1}=CM_a$	$\frac{CM_a}{CM_e}$	Ftab_*a*	*Pa*
Factor *b*	SC_b	$(b-1)$	$\frac{SC_b}{b-1}=CM_b$	$\frac{CM_b}{CM_e}$	Ftab_*b*	*Pb*
Error	SC_{Error}	$(a-1)(b-1)$	$\frac{SC_{cerror}}{(a-1)(b-1)}=CM_e$			
Total	SC_{Tot}	$N-1$				

En la tabla 4.2 llama la atención la columna titulada Promedio de los cuadrados, también denominada Cuadrados medios y regularmente se simboliza por *CM*.

La información contenida en el cuadro propone que habrá dos pruebas de hipótesis, bajo el criterio expuesto anteriormente para la *F* de Fisher, una para determinar si el efecto del factor *a* genera o no diferencias y otra para determinar si el efecto del factor *b* genera o no diferencias.

Con el propósito de mostrar de manera aplicada el procedimiento anterior, considérese el planteamiento del problema propuesto como ejemplo.

Alerta

La suma de cuadrados del factor *a* se basa en los cuadrados de las columnas, y la suma de cuadrados del factor *b* se basa en los cuadrados de los renglones.

Alerta

ANOVA para dos factores propone dos pruebas de hipótesis bajo la *F* de Fisher, una para el factor *a* y otra para el factor *b*.

Problema resuelto

Una empresa pretende innovar en el proceso de fabricación de uno de sus productos, por lo que prueba tres presentaciones diferentes del insumo principal: barra, granulado y en polvo. Ensaya cada presentación en las ocho máquinas de las que dispone, de manera que la gerencia de manufactura se interroga si existen diferencias significativas en el número de unidades producidas dependiendo del

Estadísticos e-Books & Papers

Problema resuelto *(continuación)*

tipo de presentación del insumo y del tipo de máquina. Procede a realizar un análisis de prueba de hipótesis estadísticas por ANOVA de dos factores de una muestra con un nivel de significancia del 95%.

Tabla 4.3

Factor *b* Tipo de insumo

Factor *a* Máquina

	Barra	Granulado	En polvo
Máquina 1	28	31	30
Máquina 2	26	29	31
Máquina 3	26	27	29
Máquina 4	24	27	28
Máquina 5	20	25	27
Máquina 6	22	23	28
Máquina 7	24	25	27
Máquina 8	25	26	29

Solución

Se realizan los cálculos correspondientes para la determinación de las variaciones. Los resultados se muestran en la siguiente tabla.

Tabla 4.4

	Barra	Granulado	En polvo	$\bar{Y}_j$	*Sum_Reng*	$(Sum\_Reng)^2$
M1	28.00	31.00	30.00	29.667	89.00	7 921.00
M2	26.00	29.00	31.00	28.667	86.00	7 396.00
M3	26.00	27.00	29.00	27.333	82.00	6 724.00
M4	24.00	27.00	28.00	26.333	79.00	6 241.00
M5	20.00	25.00	27.00	24.000	72.00	5 184.00
M6	22.00	23.00	28.00	24.333	73.00	5 329.00
M7	24.00	25.00	27.00	25.333	76.00	5 776.00
M8	25.00	26.00	29.00	26.667	80.00	6 400.00
$\bar{Y}_j$	24.37	26.62	28.62	$\Sigma\,(Sum\_Reng)^2 =$		50 971.00
Sum_Col	195.00	213.00	229.00			
$(Sum\_Col)^2$	38 025.00	45 369.00	52 441.00			
$\Sigma\,(Sum\_Col)^2 = 135\,835.000$						

Por tanto,

a) Promedio de promedios

$$\bar{\bar{Y}} = 26.540$$

Estadísticos e-Books & Papers

Solución *(continuación)*

b) Suma total y suma total al cuadrado

$$Sum\_Total = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij} = (28 + 31 + 30 + \ldots + 28 + 29) = 637$$

$$Sum\_Total^2 = \left(\sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij} \right)^2 = (637)^2 = 405\,769$$

c) Suma de cuadrados

$$\sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij}{}^2 = (28^2 + 31^2 + 30^2 + \ldots + 28^2 + 29^2) = 17\,081$$

d) Suma de cuadrados del factor *a*

$$SC_a = \frac{\sum_{j=1}^{b} Sum\_Col_j^2}{b \cdot n} - \frac{Sum\_Total^2}{N} = \frac{50\,971}{3 \cdot 1} - \frac{405\,769}{8 \cdot 3 \cdot 1} = 83.292$$

e) Suma de cuadrados del factor *b*

$$SC_b = \frac{\sum_{i=1}^{a} Sum\_Reng_i^2}{a \cdot n} - \frac{Sum\_Total^2}{N} = \frac{135\,835}{8 \cdot 1} - \frac{405\,769}{8 \cdot 3 \cdot 1} = 72.333$$

f) Suma de cuadrados totales

$$SC_{Tot} = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij}{}^2 - \frac{Sum_{Total}{}^2}{N} = 17\,081 - \frac{405\,769}{8 \cdot 3 \cdot 1} = 173.958$$

g) Se calcula el valor del cuadrado del error o simplemente el error.

$$SC_{Error} = SC_{Tot} - [SC_a + SC_b] = 173.958 - [83.292 + 72.333] = 18.333$$

Por lo que, con los parámetros calculados se forma el cuadro de análisis.

Tabla 4.5 Análisis de la varianza

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados
Factor *a*	83.291	$(8-1)=7$	$CM_a = \frac{83.291}{7} = 11.898$
Factor *b*	72.333	$(3-1)=2$	$CM_b = \frac{72.333}{2} = 36.167$
Error	18.333	$(8-1)(3-1)=14$	$CM_e = \frac{18.333}{14} = 1.309$
Total	173.958	$24-1=23$	

Tabla 4.6 Origen de las variaciones

Origen de las variaciones	F_{calc}	F_{tab}
Factor *a*	$\frac{CM_a}{CM_e} = \frac{11.898}{1.309} = 9.089$	Ftab_a = 2.764
Factor *b*	$\frac{CM_b}{CM_e} = \frac{36.167}{1.309} = 27.695$	Ftab_b = 3.783

Estadísticos e-Books & Papers

Solución *(continuación)*

A efecto de comprobar los cálculos anteriores, se presenta la tabla de resumen de resultados de una prueba de análisis de la varianza de dos factores con una sola muestra por grupo.

Tabla 4.7 Análisis de la varianza

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	*F*	Prob.	Valor crítico para *F*
Filas	83.291	7	11.898	9.0863	0.000272	2.7641
Columnas	72.333	2	36.166	27.618	1.382E-05	3.7388
Error	18.333	14	1.309			
Total	**173.958**	**23**				

El análisis de los resultados, en relación con la prueba de hipótesis estadística, muestra que tanto en el factor *a* como en el factor *b* se cumplen las hipótesis H_1 (hay diferencia), lo que quiere decir que los factores generan un efecto sobre el número de unidades producidas.

Alerta

El ANOVA para dos factores se puede llevar a cabo a través de Excel mediante la prueba de análisis de la varianza de dos factores con una sola muestra por grupo.

4.8 Estudios comparativos basados en ANOVA de dos factores de varias muestras por grupo

Este tipo de estudios sirven para evaluar el comportamiento individual y en conjunto de dos factores sobre una variable dependiente cuantitativa, pero considerando que los tratamientos pueden ser estudiados a través de varias muestras.

Tal como se expuso, el análisis permite observar la incidencia o efectos de cada variable individual, así como la interacción entre ambas; es decir, se estudian tres efectos.

Para precisar considérese el caso que se cuente con *m* muestras con tamaño *n* dispuestas de forma matricial, donde uno de los factores en estudio, denotado por la letra *a*, estará definido por un número de ensayos que componen la muestra, mientras que el segundo factor, denominado *b*, estará referido por las columnas.

Tabla 4.8

Factor *a*	Factor *b*	Barra	Granulado	En polvo
Cantidad *X*		28.000	31.000	30.000
		26.000	29.000	31.000
		26.000	27.000	29.000
		24.000	27.000	28.000
Cantidad *Y*		20.000	25.000	27.000
		22.000	23.000	28.000
		24.000	25.000	27.000
		25.000	26.000	29.000

Alerta

El ANOVA de dos factores con varias muestras por grupo refiere que el factor *a* cuenta con varios ensayos.

Estadísticos e-Books & Papers

De manera práctica, el procedimiento de cálculo se expone al tener como base el ejemplo del procedimiento anterior y en el siguiente planteamiento: la empresa está interesada en analizar el impacto de la cantidad de insumo según la presentación en el número de piezas producidas, así como en saber si de manera combinada también inciden.

Obsérvese que se ensayan dos cantidades diferentes en las que la relación renglón-columna cuenta con un número de cuatro ensayos, es decir, el tamaño muestral es de cuatro elementos.

De manera que el procedimiento de cálculo bajo el enfoque de ANOVA de dos factores de varias muestras requiere del cálculo de las sumas de los cuadrados de los efectos de cada factor, así como en forma conjunta, tal como se expone a continuación.

a) Para el factor *a* (renglones):

$$SC_a = \frac{\Sigma_{j1} Sum\_Reng^2}{b \cdot n} - \frac{Sum\_Total^2}{N}$$

b) Para el factor *b* (columnas):

$$SC_b = \frac{\Sigma Sum\_Col^2}{a \cdot n} - \frac{Sum\_Total^2}{N}$$

c) Suma de cuadrados de ambos factores:

$$SC_{ab} = \frac{\Sigma Am^2}{n} - \left(\frac{Sum\_Total^{\ 2}}{N} + SC_a + SC_b \right)$$

Am es el acumulado por muestra.

d) Se calcula el valor de los cuadrados totales:

$$SC_{Tot} = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij}^{\ 2} - \frac{Sum\_Total^{\ 2}}{N}$$

e) Se calcula el valor de los cuadrados del error o simplemente el error:

$$SC_{Error} = SC_{Tot} - [SC_a + SC_b + SC_{ab}]$$

Con base en los parámetros anteriores se estructura el cuadro de análisis.

Tabla 4.9 Análisis de la varianza

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados
Factor *a*	SC_a	$(a-1)$	$\frac{SC_a}{a-1} = CM_a$
Factor *b*	SC_b	$(b-1)$	$\frac{SC_b}{b-1} = CM_b$
Interacción *ab*	SC_{ab}	$(a-1)(b-1)$	$\frac{SC_{ab}}{(a-1)(b-1)} = CM_{ab}$
Error	SC_{Error}	$N-(ab)$	$\frac{SC_{cerror}}{N-(ab)} = CM_e$
Total	SC_{Tot}	$N-1$	

561
Estadísticos e-Books & Papers

Tabla 4.10

Origen de las variaciones	F_{calc}	F_{tab}	Prob.
Factor *a*	$\frac{CM_a}{CM_e}$	Ftab_*a*	*Pa*
Factor *b*	$\frac{CM_b}{CM_e}$	Ftab_*b*	*Pb*
Interacción *ab*	$\frac{CM_{ab}}{CM_e}$	Ftab_*b*	*Pab*
Error			
Total			

La información contenida en el cuadro propone que habrá tres pruebas de hipótesis, bajo el criterio expuesto para la *F* de Fisher: una para determinar si el efecto del factor *a* genera o no diferencias; una segunda para determinar si el efecto del factor *b* origina o no diferencias; y una tercera para determinar si la interacción de los factores genera o no diferencias.

Con el propósito de mostrar de manera aplicada el procedimiento anterior, considérese el planteamiento del problema propuesto como ejemplo.

Alerta

El ANOVA para dos factores de varias muestras propone tres pruebas de hipótesis bajo la *F* de Fisher: una primera para el factor *a*; una segunda para el factor *b*; y una tercera para la interacción de los dos factores.

Problema resuelto

La empresa está interesada en analizar el impacto de la cantidad de insumo dependiendo de la presentación en el número de piezas producidas, así como en saber si de manera combinada también inciden.

Para ello realiza un análisis de prueba de hipótesis estadísticas por ANOVA de dos factores de una muestra con un nivel de significancia del 95%.

Tabla 4.11

Factor *b*

Factor *a*	Barra	Granulado	En polvo
Cantidad *X*	28.000	31.000	30.000
	26.000	29.000	31.000
	26.000	27.000	29.000
	24.000	27.000	28.000
Cantidad *Y*	20.000	25.000	27.000
	22.000	23.000	28.000
	24.000	25.000	27.000
	25.000	26.000	29.000

Estadísticos e-Books & Papers

Solución

Se realizan los cálculos correspondientes para la determinación de las variaciones. Los resultados se fundamentan en los obtenidos del procedimiento anterior complementándolos como se muestra.

Se calculan los valores que se muestran en la siguiente tabla.

Tabla 4.12

	Barra		Granulado		En polvo		Suma de renglones
Cantidad *X*	28.000 26.000 26.000 24.000	Am_1 104.000	31.000 29.000 27.000 27.000	Am_2 114.000	30.000 31.000 29.000 28.000	Am_3 118.000	336.000
Cantidad *Y*	20.000 22.000 24.000 25.000	Am_4 91.000	25.000 23.000 25.000 26.000	Am_5 99.000	27.000 28.000 27.000 29.000	Am_6 111.000	301.000
Suma de columnas		195.000		213.000		229.000	Total 637
Cantidad *X*	$\bar{X}_{CX}$	26.000		28.500		29.500	
Cantidad *Y*	$\bar{X}_{CY}$	22.750		24.750		27.750	
	$\bar{x}$	26.542					

$$\Sigma(Sum\_Col)^2 = 195^2 + 213^2 + 229^2 = 135\,835$$

$$\Sigma(Sum\_Reng)^2 = 336^2 + 301^2 = 203\,497$$

$$\Sigma(Am)^2 = 104^2 + 114^2 + 118^2 + 91^2 + 99^2 + 111^2 = 68\,139$$

Considérese que el factor *a* está conformado por dos grupos, mientras que el factor *b*, por tres grupos, por tanto:

$$a = 2 \qquad b = 3$$

Adicionalmente, la relación renglón-columna expone que se compone de cuatro elementos, por lo que el tamaño muestral (*n*) es 4.

Por consiguiente,

a) Suma total y suma total al cuadrado:

$$Sum\_Total = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij} = (28 + 31 + 30 + \ldots + 28 + 29) = 637$$

$$Sum\_Total^2 = \left(\sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij} \right)^2 = (637)^2 = 405\,769$$

Estadísticos e-Books & Papers

Solución *(continuación)*

b) Suma de cuadrados del factor *a*:

$$SC_a = \frac{\Sigma\ Sum\_Reng^2}{b \cdot n} - \frac{Sum\_Total^{\,2}}{N} = \frac{203\,497}{3 \cdot 4} + \frac{405\,769}{2 \cdot 3 \cdot 4} = 51.0417$$

c) Suma de cuadrados del factor *b*:

$$SC_b = \frac{\Sigma\ Sum\_Col^2}{a \cdot n} - \frac{Sum\_Total^{\,2}}{N} = \frac{135\,835}{2 \cdot 4} + \frac{405\,769}{2 \cdot 3 \cdot 4} = 72.333$$

d) Suma de cuadrados de la interacción de factores:

$$SC_{ab} = \frac{\Sigma Am^2}{n} - \left(\frac{Sum\_Total^{\,2}}{N} + SC_a + SC_b\right) =$$

$$= \frac{68\,139}{4} - \left(\frac{405\,769}{24} + 51.0417 + 72.333\right) = 4.333$$

e) Suma de cuadrados totales:

$$SC_{Tot} = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^{\,2} = \frac{Sum_{Total}^{\,2}}{N} = 17\,081 - \frac{405\,769}{8 \cdot 3 \cdot 1} = 173.958$$

f) Se calcula el valor del cuadrado del error o simplemente el error.

$$SC_{Error} = SC_{Tot} - [SC_a + SC_b + SC_{ab}]$$

$$SC_{Error} = SC_{Tot} - [SC_a + SC_{b\,+}\, SC_{ab}]$$

$$= 173.958 - [51.0417 + 72.333 + 4.333] = 46.275$$

Tabla 4.13 Análisis de la varianza

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados
Factor *a*	51.0417	$(a-1) = 2-1 = 1$	$CM_a = \frac{51.0417}{1} = 51.0417$
Factor *b*	72.333	$(b-1) = 3-1 = 2$	$CM_b = \frac{72.333}{2} = 36.1666$
Interacción *ab*	4.3333	$(a-1)(b-1) =$ $(2-1)(3-1) = 2$	$CM_{ab} = \frac{4.333}{2} = 2.1666$
Error	46.275	$N-(a\,b) =$ $24-(2\cdot 3) = 18$	$CMe = \frac{46.275}{18} = 2.5708$
Total	173.958	$N-1 = 24-1 = 23$	

Estadísticos e-Books & Papers

Alerta

El ANOVA para dos factores se puede llevar a cabo a través de Excel mediante la prueba de análisis de la varianza de dos factores con varias muestras por grupo.

Solución *(continuación)*

Tabla 4.14

Origen de las variaciones	F_{calc}	F_{tab}
Factor *a*	$\frac{CM_a}{CM_e}=19.8544$	4.4139
Factor *b*	$\frac{CM_b}{CM_e}=14.0682$	3.5546
Interacción *ab*	$\frac{CM_{ab}}{CM_e}=0.8427$	3.5546
Error		
Total		

De primera instancia se interpreta que la cantidad de material aplicado al proceso de fabricación sí genera diferencias en el número de unidades producidas.

En cuanto al tipo de presentación del insumo, se puede señalar que éste también genera diferencias en el número de unidades producidas. Sin embargo, la interacción de los factores expone que no deberían generar diferencias.

Asimismo, se muestra la tabla generada por Excel para el análisis de la varianza de dos factores con varias muestras por grupo a efecto de corroborar los resultados obtenidos.

Tabla 4.15 Análisis de la varianza

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	*F*	Prob.	Valor crítico para *F*
Muestra	51.042	1	51.042	19.865	0.000	4.414
Columnas	72.333	2	36.167	14.076	0.000	3.555
Interacción	4.333	2	2.167	0.843	0.447	3.555
Dentro del grupo	46.250	18	2.569			
Total	173.958	23				

4.9 Discriminantes para pruebas de hipótesis basadas en ANOVA

Las pruebas de hipótesis por ANOVA pueden establecerse de orden general, ya que tan solo señalan si se aprueba o refuta la hipótesis nula. Sin embargo, cuando se cuenta con más de dos grupos y se aprueba la hipótesis alternativa surge la interrogante: ¿entre qué grupos existe la diferencia? Un proceso lógico sería realizar pruebas de ANOVA por pares de grupos a efecto de determinar los grupos que difieren entre sí, pero esta alternativa de cálculo puede resultar larga y tediosa, por lo que en su momento se crearon procesos de discriminación fundamentados en los cuadrados promedio (varianzas) y la distribución *t*.

Para propósitos de los contenidos de esta obra se hará referencia a dos discriminantes: el factor de Diferencia Mínima Significativa (*LSD*, *Least Significant Difference*), así como la prueba de rangos múltiples de Duncan.

Alerta

Utilizar muestras pequeñas expone que el valor de la varianza poblacional se considera desconocida, pero puede ser sustituida por el promedio de los cuadrados del error.

Discriminante *LSD* para prueba de hipótesis de un factor con más de dos grupos

El análisis por el discriminante *LSD* se fundamenta en el supuesto de que la desviación estándar poblacional (σ) es desconocida, pero puede sustituirse por $\sqrt{CM_e}$, al considerar los resultados obtenidos a

Estadísticos e-Books & Papers

través del cálculo del análisis de la varianza correspondiente, de manera que se puede complementar con la aplicación de la distribución *t* de Student considerando un factor de escala $\sqrt{\frac{CM_e}{n}}$.

Se entiende que el valor del CM_e corresponde al promedio de los cuadrados dentro de grupos.

De manera que cuando el análisis contempla grupos con el mismo tamaño muestral *n*, el valor del discriminante *LSD* se determina mediante la fórmula

$$LSD = t_{\frac{\alpha}{2},N-N_m}\sqrt{\frac{2CM_e}{n}}$$

Pero cuando los grupos cuentan con tamaños muestrales diferentes, el valor del discriminante se obtiene al aplicar la fórmula

$$LSD = t_{\frac{\alpha}{2},N-N_m}\sqrt{CM_e\left(\frac{1}{n_i}+\frac{1}{nj}\right)}$$

Donde

α = valor del error

$N - Nm$ = los grados de libertad del error (variación dentro de grupos)

n_i = tamaño muestral del primer grupo

n_j = tamaño muestral del segundo grupo

El criterio de discriminación es el siguiente:

Si $\left|\overline{y}_i + \overline{y}_j\right| > LSD$ se concluye que existe diferencia entre las medias poblacionales μ_i y μ_j.

Para ejemplificar el proceso de cálculo, considérese el ejemplo sobre la resina de los cartuchos de escopeta, pero ténganse en cuenta las tres resinas.

Problema resuelto

Realiza un análisis comparativo simple de ANOVA para determinar si existe diferencia en los cartuchos dependiendo del tipo de resina del cuerpo si se considera el peso, en gramos, de seis cartuchos por muestra. Considera un nivel de confianza del 95%.

Tabla 4.16

Resina A (RA)	Resina A-1 (RA1)	Resina A-2 (RA2)
32.30	31.89	32.10
31.90	32.10	32.00
31.45	32.34	31.90
32.80	32.40	32.60
33.01	32.30	32.66
32.78	31.99	32.39

Estadísticos e-Books & Papers

Alerta

El análisis por *LSD* de un factor se puede realizar tanto en el sentido vertical como en el horizontal, dependiendo de las necesidades de análisis.

Solución

La solución se logra mediante Excel: análisis de la varianza de un factor.

Tabla 4.17 Análisis de la varianza de un factor

Resumen				
Grupos	Cuenta	Suma	Promedio	Varianza
RA	6	194.24	32.3733	0.3669
RA1	6	193.02	32.1700	0.0429
RA2	6	193.63	32.2717	0.1029

Análisis de la varianza						
Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	*F*	Prob.	Valor crítico para *F*
Entre grupos	0.1240	2	0.0620	0.3629	0.7016	3.6823
Dentro de los grupos	2.5636	15	0.1709			
Total	2.6876	17				

Donde se identifica que el valor del $CM_e = 0.1709$.

$$t_{\frac{\alpha}{2},N-N_m} = t_{\frac{0.05}{2}(18-3)} = t_{0.025,15} = 2.131$$

Por lo que el valor del discriminante es:

$$LSD = t_{\frac{\alpha}{2},N-N_m}\sqrt{\frac{2CM_e}{n}} = (2.131)\sqrt{\frac{(2)(0.1709)}{6}} = 0.2387$$

Se realiza el proceso de discriminación como sigue:

$$\bar{y}_{RA} = 32.3733 \qquad \bar{y}_{RA1} = 32.1700 \qquad \bar{y}_{RA2} = 32.2717$$

$$\left|\bar{\bar{Y}}_{RA} - \bar{\bar{Y}}_{RA1}\right| = |32.3733 - 32.1700| = 0.2033 < LSD \quad \text{no hay diferencia}$$

$$\left|\bar{\bar{Y}}_{RA} - \bar{\bar{Y}}_{RA2}\right| = |32.3733 - 32.2717| = 0.1016 < LSD \quad \text{no hay diferencia}$$

$$\left|\bar{\bar{Y}}_{RA1} - \bar{\bar{Y}}_{RA1}\right| = |32.1700 - 32.2717| = 0.1017 < LSD \quad \text{no hay diferencia}$$

Se concluye que los tipos de resina no ofrecen diferencia entre los grupos.

Discriminante *LSD* para prueba de hipótesis de dos factores con más de dos grupos y varias muestras por grupo

Bajo este criterio se puede señalar que tanto un factor *a* puede generar diferencias entre las medias poblacionales como las puede generar un factor *b*, por lo que se tendrá que calcular un determinante *LSD* para *a* y un discriminante para el factor *b*.

Estadísticos e-Books & Papers

De manera concreta, el procedimiento expone comparar las medias muestrales de los grupos por pares a efecto de precisar entre qué grupos existen las diferencias.

Para el factor *a* el valor del discriminante se calcula mediante la fórmula

$$LSD_a = t_{\frac{\alpha}{2},ab(n-1)}\sqrt{CM_e\left(\frac{1}{n_{ai}}+\frac{1}{n_{aj}}\right)}$$

Asimismo, para el factor *b* el valor del discriminante se calcula mediante la fórmula

$$LSD_b = t_{\frac{\alpha}{2},ab(n-1)}\sqrt{CM_e\left(\frac{1}{n_{bi}}+\frac{1}{n_{bj}}\right)}$$

Donde

α = valor del error

$ab(n-1)$ = grados de libertad del error (variación dentro de grupos)

n_{ai} o n_{bi} = tamaño muestral del primer grupo

n_{aj} o n_{bj} = tamaño muestral del segundo grupo

Se puede generar el discriminante para la interacción de los factores:

$$LSD_{ab} = t_{\frac{\alpha}{2},ab(n-1)}\sqrt{CM_e\left(\frac{1}{n}+\frac{1}{n}\right)}$$

donde *n* es el número de observaciones en la relación factor *a*–factor *b*.

El criterio de discriminación para cualquiera de los factores es el siguiente.

Si $\left|\overline{y}_i - \overline{y}_j\right| > LSD$ se concluye que existe diferencia entre las medias poblacionales μ_i y μ_j. Debe apuntarse que en el análisis de la interacción las medias corresponden al promedio basado en los AG.

Nótese que en las fórmulas se consideran los tamaños muestrales de los grupos en análisis de manera que puedan compararse estos en caso de que el tamaño muestral no sea homogéneo.

Para ejemplificar lo anterior, considérese de nuevo el problema de los cartuchos del punto anterior pero bajo las condiciones que se exponen.

Problema resuelto

Realiza un análisis comparativo simple para determinar si existe diferencia en el peso de los cartuchos dependiendo del tipo de resina del cuerpo y de pólvora. Se consideran muestras de tres elementos y un nivel de confianza del 95%.

Tabla 4.18

	Resina A (RA)	Resina A-1 (RA1)	Resina A-2 (RA2)
Pólvora 1 (*P1*)	32.30	31.89	32.10
	31.90	32.10	32.00
	31.45	32.34	31.90
Pólvora 2 (*P2*)	32.80	32.40	32.60
	33.01	32.30	32.66
	32.78	31.99	32.39

Estadísticos e-Books & Papers

Solución

La solución se logra mediante Excel: análisis de la varianza de dos factores de varias muestras por grupo.

Tabla 4.19

	Resina RA		Resina RA1		Resina RA2	
Pólvora *P1*	32.30	$Am_{P1\text{-}RA}$	31.89	$Am_{P1\text{-}RA1}$	32.10	$Am_{P1\text{-}RA2}$
	31.90	31.88	32.10	32.11	31.96	31.98
	31.45		32.34		31.90	
Pólvora *P2*	32.80	$Am_{P2\text{-}RA}$	32.40	$Am_{P2\text{-}RA1}$	32.60	$Am_{P2\text{-}RA2}$
	33.01	32.86	32.30	32.23	32.72	32.57
	32.78		31.99		32.39	

Tabla 4.20 Análisis de la varianza de dos factores con varias muestras por grupo

Resumen	RA	RA1	RA2	Total
Cuenta	3	3	3	9
Suma	95.6500	96.3300	95.9900	287.9700
Promedio	31.8833	32.1100	31.9967	31.9967
Varianza	0.1808	0.0507	0.0100	0.0700
Cuenta	3	3	3	9
Suma	98.5900	96.6900	97.6400	292.9200
Promedio	32.8633	32.2300	32.5467	32.5467
Varianza	0.0162	0.0457	0.0204	0.0958
Total				
Cuenta	6	6	6	
Suma	194.2400	193.0200	193.6300	
Promedio	32.3733	32.1700	32.2717	
Varianza	0.3669	0.0429	0.1029	

Tabla 4.21 Análisis de la varianza

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	*F*	Prob.	Valor crítico para *F*
Muestra	1.3612	1	1.3612	25.2213	0.0003	4.7472
Columnas	0.1240	2	0.0620	1.1490	0.3495	3.8853
Interacción	0.5547	2	0.2774	5.1388	0.0244	3.8853
Dentro del grupo	0.6477	12	0.0540			

Estadísticos e-Books & Papers

Solución *(continuación)*

Por tanto, el valor de los discriminantes para los dos factores es el siguiente

$$t_{\frac{\alpha}{2},ab(n-1)} = t_{\frac{0.5}{2},(2)(3)(3-1)} = t_{0.025,12} = 2.179$$

$$LSD_a = t_{\frac{\alpha}{2},ab(n-1)}\sqrt{CM_e\left(\frac{1}{n_{ai}}+\frac{1}{n_{aj}}\right)} = (2.179)\sqrt{(0.0540)\left(\frac{1}{9}+\frac{1}{9}\right)} = 0.2387$$

$$\left|\overline{y_{P1}} - \overline{y_{P2}}\right| = |31.9967 - 32.5467| = 0.55 > LSD \quad \text{existe diferencia}$$

Asimismo, para el factor b el valor del discriminante se calcula mediante la fórmula

$$LSD_b = t_{\frac{\alpha}{2},ab(n-1)}\sqrt{CM_e\left(\frac{1}{n_{bi}}+\frac{1}{n_{bj}}\right)} = (2.179)\sqrt{(0.0540)\left(\frac{1}{6}+\frac{1}{6}\right)} = 0.2923$$

$$\overline{y}_{RA} = 32.3733 \qquad \overline{y}_{RA1} = 32.1700 \qquad \overline{y}_{RA2} = 32.2717$$

$$\left|\overline{\overline{Y}}_{RA} - \overline{\overline{Y}}_{RA1}\right| = |32.3733 - 32.1700| = 0.2033 < LSD \quad \text{no hay diferencia}$$

$$\left|\overline{\overline{Y}}_{RA} - \overline{\overline{Y}}_{RA2}\right| = |32.3733 - 32.2717| = 0.1016 < LSD \quad \text{no hay diferencia}$$

$$\left|\overline{\overline{Y}}_{RA1} - \overline{\overline{Y}}_{RA1}\right| = |32.1700 - 32.2717| = 0.1017 < LSD \quad \text{no hay diferencia}$$

Para el análisis *LSD* de la interacción se determinan los valores de los promedios por cada conjunto de observaciones ($n = 3$).

$$\overline{Y}_{P1-RA} = \frac{31.88}{3} = 10.63 \qquad \overline{Y}_{P2-RA} = \frac{32.86}{3} = 10.95$$

$$\overline{Y}_{P1-RA1} = \frac{32.11}{3} = 10.70 \qquad \overline{Y}_{P2-RA1} = \frac{32.23}{3} = 10.74$$

$$\overline{Y}_{P1-RA2} = \frac{31.98}{3} = 10.66 \qquad \overline{Y}_{P2-RA2} = \frac{32.57}{3} = 10.86$$

El valor del discriminante de la interacción es

$$LSD_{ab} = t_{\frac{\alpha}{2},ab(n-1)}\sqrt{CM_e\left(\frac{1}{n}+\frac{1}{n}\right)} = (2.179)\sqrt{(0.0540)\left(\frac{1}{3}+\frac{1}{3}\right)} = 0.4134$$

Al realizar el proceso de discriminación por relación de factores, de manera básica son los siguientes:

$$\left|\overline{Y}_{P1-RA} - \overline{Y}_{P2-RA}\right| = |10.63 - 10.95| = 0.32 < LSD \quad \text{no hay diferencia}$$

$$\left|\overline{Y}_{P1-RA1} - \overline{Y}_{P2-RA1}\right| = |10.66 - 10.74| = 0.08 < LSD \quad \text{no hay diferencia}$$

$$\left|\overline{Y}_{P1-RA2} - \overline{Y}_{P2-RA2}\right| = |10.66 - 10.86| = 0.2 < LSD \quad \text{no hay diferencia}$$

Estadísticos e-Books & Papers

Solución *(continuación)*

Pueden llevarse a cabo otras tantas discriminaciones dependiendo de los requerimientos de análisis que se consideren oportunos.

La conclusión es que los factores considerados para el experimento no generan una diferencia significativa en el diseño del cartucho; en otras palabras, se pueden fabricar indistintamente con cualquiera de ellos.

4.10 Método de discriminación de la *R* de Duncan

Alerta

El análisis por *R* de Duncan aplica únicamente para análisis de un solo factor con muestras del mismo tamaño.

El método creado en 1955 permite comparar pares de medias poblacionales y se conoce como la Prueba de Rangos Múltiples de Duncan para análisis de un solo factor con muestras del mismo tamaño.

El método se fundamenta en el promedio del error estándar definido por:

$$S\overline{y}_i = \sqrt{\frac{CME}{n}}$$

De manera que se procede a encontrar el valor de la *R* en las tablas de los rangos significativos de Duncan:

$$Rp = r_\alpha\ (p, f) S\overline{y}_i$$

Donde

r_α = rangos significativos de Duncan

f = grados de libertad del error

P = valores (2, 3, 4,… a)

Para realizar el análisis se debe hacer lo siguiente.

1. Calcular el valor de $S\overline{y}_i$.
2. Calcular los valores de $Rp(R_2, R_3,\ldots Ra)$.
3. Ordenar las medias muestrales de menor a mayor.
4. Encontrar las diferencias entre las medias empezando por la de valor más alto contra las restantes en orden descendente.
5. Comparar las diferencias contra los valores de Rp de manera descendente. El criterio de evaluación es:

   $\overline{y}_i - \overline{y}_j > Rp$ existe diferencia

   Para mostrar el procedimiento expuesto considérese el problema de los cartuchos.

Problema resuelto

La fábrica de armas experimenta con otros tipos de resina modificados para generar estabilidad e impermeabilidad a los cartuchos de uno de sus tipos de escopeta. Realiza un análisis comparativo simple, basado en el método de la *R* de Duncan, para determinar si existe diferencia entre los cartuchos dependiendo del tipo de resina del cuerpo. El nivel de confianza es de 95%.

571
Estadísticos e-Books & Papers

Problema resuelto *(continuación)*

Tabla 4.22

Resina A (RA)	Resina A-1 (RA1)	Resina A-2 (RA2)	Resina A-3 (RA3)	Resina A-4 (RA4)
32.400	33.450	32.260	32.703	32.925
32.500	32.980	32.340	32.607	32.740
31.950	32.340	32.200	32.163	32.145
32.800	33.150	31.980	32.643	32.975

Solución

Con base en los datos se determinan las medias de cada grupo.

Tabla 4.23

Resina A (RA)	Resina A-1 (RA1)	Resina A-2 (RA2)	Resina A-3 (RA3)	Resina A-4 (RA4)
32.400	33.450	33.440	33.097	32.925
32.500	32.980	32.420	32.633	32.740
31.950	32.340	32.960	32.417	32.145
32.800	33.150	33.110	33.020	32.975
32.413	32.980	32.983	32.792	32.696

Se procede a determinar el valor del CM_e por medio de la rutina de Excel de análisis de la varianza para un solo factor.

Tabla 4.24 Análisis de la varianza

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	*F*	Prob.	Valor crítico para *F*
Entre grupos	0.8918	4	0.2229	1.4411	0.2689	3.0556
Dentro de los grupos	2.3204	15	0.1547			
Total	3.2122	19				

Calculamos el valor promedio del error estándar

$$S\bar{y}_i = \sqrt{\frac{CM_e}{n}} = \sqrt{\frac{0.1547}{4}} = 0.1967$$

Determinamos los valores de los rangos significativos de Duncan utilizando las tablas

$$r_{0.05}(2,15) = 3.01$$
$$r_{0.05}(3,15) = 3.16$$
$$r_{0.05}(4,15) = 3.25$$
$$r_{0.05}(5,15) = 3.31$$

Estadísticos e-Books & Papers

Solución *(continuación)*

Por lo que

$R_2 = r_{0.05}(2,15)\, S\overline{y}_i = (3.01)(0.1967) = 0.5961$
$R_3 = r_{0.05}(3,15)\, S\overline{y}_i = (3.16)(0.1967) = 0.6216$
$R_4 = r_{0.05}(4,15)\, S\overline{y}_i = (3.25)(0.1967) = 0.6393$
$R_5 = r_{0.05}(5,15)\, S\overline{y}_i = (3.31)(0.1967) = 0.6511$
$Rp = r\alpha\,(p, f)\, S\overline{y}_i$

Al ordenar las medias de menor a mayor

$\overline{Y}_{RA} = 32.413$
$\overline{Y}_{RA4} = 32.696$
$\overline{Y}_{RA3} = 32.786$
$\overline{Y}_{RA2} = 32.965$
$\overline{Y}_{RA1} = 32.980$

*RA*1 *v.s. RA*: 32.980 – 32.413 = 0.5670 < 0.6511 (R_5) no existe diferencia
*RA*1 *v.s. RA*4: 32.980 – 32.696 = 0.284 < 0.6393 (R_4) no existe diferencia
*RA*1 *v.s. RA*3: 32.980 – 32.786 = 0.194 < 0.6216 (R_3) no existe diferencia
*RA*1 *v.s. RA*2: 32.980 – 32.965 = 0.015 < 0.5961 (R_2) no existe diferencia

*RA*2 *v.s. RA*: 32.965 – 32.413 = 0.5520 < 0.6393 (R_4) no existe diferencia
*RA*2 *v.s. RA*4: 32.965 – 32.696 = 0.2690 < 0.6216 (R_3) no existe diferencia
*RA*2 *v.s. RA*3: 32.965 – 32.786 = 0.1790 < 0.5961 (R_2) no existe diferencia

*RA*3 *v.s. RA*4: 32.796 – 32.696 = 0.1000 < 0.6216 (R_3) no existe diferencia
*RA*2 *v.s. RA*: 32.965 – 32.413 = 0.5520 < 0.5961 (R_2) no existe diferencia

*RA*4 *v.s. RA*: 32.696 – 32.413 = 0.2830 < 0.5961 (R_2) no existe diferencia

La conclusión es que las resinas para el cuerpo del cartucho no proponen ninguna diferencia en su peso.

Estadísticos e-Books & Papers

4.1 Una empresa fabricante de llantas especiales para vehículos de la industria de la construcción cuenta con cuatro plantas donde se fabrica el mismo modelo de llanta con los niveles de producción que se muestran.

Tabla 4.25

Planta A	Planta B	Planta C	Planta D
26	37	41	36
38	34	38	32
32	38	30	30
36	38	26	34
34	44	22	36

En cada planta se reproduce de igual manera el proceso de manufactura y la gerencia de producción desea saber si existen diferencias en la producción de cada planta a través de un estudio comparativo simple si se considera un error del 1%.

4.2 Resuelve el problema anterior con un error al 5%.

4.3 Una planta fabricante de un filamento especial para lámparas incandescentes labora dos turnos de ocho horas (matutino y vespertino) con los niveles de producción en metros que se citan en la siguiente tabla.

Tabla 4.26

Matutino	Vespertino
22.420	23.510
18.360	20.620
21.460	26.470
19.200	19.750
23.400	20.300
27.380	17.840
23.460	26.340

La gerencia de control de calidad está interesada en conocer si hay diferencia en la producción entre los turnos, por lo que propone efectuar un estudio comparativo simple con un nivel de confianza de 90%.

4.4 Resuelve el problema anterior considerando un error del 5%.

4.5 Una empresa empacadora de camarón café para exportación cuenta con tres líneas de empacado con los siguientes volúmenes de producción de cajas por hora.

Tabla 4.27

Línea 1	Línea 2	Línea 3
51	49	54
50	53	54
55	48	53
48	50	52

Eres analista de una empresa y el dueño te encomienda realizar un análisis por estudios comparativos simples para establecer si existe o no diferencia en la producción de cada línea al 95% de confianza.

4.6 Una empresa fabricante de bolsas industriales para agua pone a prueba cuatro de sus diseños inyectando aire a presión (KPa) hasta que estallen.

Tabla 4.28

Tipo 1	Tipo 2	Tipo 3	Tipo 4
2.4	2.5	2.5	2.7
2.3	2.5	2.5	2.6
2.6	2.4	2.6	2.8
2.5	2.3	2.7	2.6
2.7	2.6	2.7	2.6
2.3	2.5	2.5	2.5

El fabricante desea saber si existe diferencia en la presión máxima de las bolsas dependiendo de su tipo. Se propone realizar un análisis comparativo simple con un nivel de confianza del 90%.

4.7 Analiza el problema anterior, considerando un nivel de confianza del 95%.

4.8 Un laboratorio realiza pruebas sobre un nuevo tipo de insulina inyectable para pacientes diabéticos insulinodependientes. Se seleccionaron 21 pacientes voluntarios con base en sus características físicas y expediente clínico. Se repartieron al azar en tres grupos; a cada grupo se le fijó una dieta igual pero con diferentes dosis del medicamento. Los investigadores realizan una prueba de hipótesis estadísticas por estudios comparativos simples para determinar si existe diferencia entre las dosis aplicadas a cada grupo al considerar un error del 5%.

Tabla 4.29

Dosis 1	Dosis 2	Dosis 3
115	102	98
108	105	98
120	100	97
110	98	100
109	95	95
108	100	102
112	104	105

4.9 Una fábrica de refacciones para autos prueba tres nuevas aleaciones para las vieletas de dirección de un auto económico popular. Los resultados de las pruebas de dureza en unidades Rockwell son las que se muestran en la tabla. La gerencia de control de calidad está interesada en saber si existen diferencias en los materia-

Estadísticos e-Books & Papers

les propuestos para la refacción, por lo que realiza una prueba de hipótesis estadísticas para estudios comparativos simples al 99% de confianza.

Tabla 4.30

Aleación 1	Aleación 2	Aleación 3
26.0	29.0	29.4
26.5	28.9	29.3
27.5	28.5	29.0
27.0	28.6	29.3

4.10 Para optimizar el tiempo de acceso a la información, una fábrica de discos duros portátiles de 500 GB realiza pruebas en los tres modelos que se encuentran en desarrollo en computadoras con el mismo procesador y obtiene los siguientes tiempos en milisegundos.

Tabla 4.31

Disco 1	Disco 2	Disco 3
8.2	8.4	8.8
8.0	8.3	8.6
8.2	8.5	8.9
8.3	8.4	8.7
8.1	8.6	8.6
8.1	8.4	8.6
8.2	8.5	8.5
8.2	8.5	8.6

Para determinar si existen diferencias entre los modelos, se pone en práctica un proceso de análisis de pruebas estadísticas simples por estudios comparativos simples al 99% de confianza.

4.11 Con base en el planteamiento del problema 4.1, establece si existen diferencias entre los grupos propuestos a través de un análisis de la varianza de un solo factor y corrobora los resultados de los cálculos mediante la rutina de Excel.

4.12 Argumenta, ¿coinciden los resultados obtenidos por el estudio comparativo simple con el de ANOVA de un solo factor en el planteamiento del problema 4.1?

4.13 Con fundamento en el planteamiento del problema 4.3 establece si existen diferencias entre los grupos propuestos a través de un análisis de la varianza de un solo factor. Corrobora los resultados de los cálculos mediante la rutina de Excel; considera un nivel de confianza del 99%.

T **4.14** Con base en el planteamiento del problema 4.5, establece si existen diferencias entre los grupos propuestos a través de un análisis de la varianza de un solo factor. Corrobora los resultados de los cálculos mediante Excel; considera niveles de confianza del 90% y 95%.

4.15 Con fundamento en el planteamiento del problema 4.6, establece si hay diferencias entre los grupos propuestos a través de un análisis de la varianza de un solo factor. Corrobora los resultados de los cálculos mediante la rutina de Excel; considera un nivel de confianza del 99%.

4.16 Una empresa de concretos experimenta con un concreto basado en tres nuevos tipos de cemento: puzolánico I, puzolánico II y puzolánico III. Se llevan a cabo pruebas de resistencia en una mezcla de concreto de 250 kg/cm^2. Realiza una prueba de ANOVA de un factor de una muestra para comparar la resistencia entre las mezclas de concreto y establece si hay diferencia en la resistencia del concreto de acuerdo con el tipo de cemento utilizado. Considera un nivel de confianza del 95%.

Tabla 4.32

Resistencia en concreto con cemento Puzolánico I	Resistencia en concreto con cemento Puzolánico II	Resistencia en concreto con cemento Puzolánico III
258	248	254
263	252	256
256	246	257
256	248	255
262	247	251
262	246	253
256	250	253
258	250	254
256	251	257
254	250	256
255	252	255
256	255	254

4.17 Con base en los resultados del problema anterior, realiza un proceso de discriminación por *LSD* de un factor.

4.18 Una empresa dedicada a la fabricación de vidrio desea extender la variedad de sus productos de seguridad. Quiere lanzar al mercado un nuevo producto al cual le han añadido resinas para variar la resistencia del vidrio dependiendo de la cantidad que se añada. La empresa desea saber si existe variación en la resistencia del vidrio (kg/cm^2) entre una nueva resina y la actual al realizar un análisis por discriminante *LSD*. En la tabla que sigue se muestra la comparación de las resistencias del vidrio con la resina actual y la nueva.

Tabla 4.33

Resistencia del vidrio Resina actual	Resistencia del vidrio Resina nueva
85	96
80	87
75	84
76	72
50	60

Estadísticos e-Books & Papers

4.19 Con base en el planteamiento del problema 4.8, haz una prueba de discriminante *LSD* de un factor con el fin de establecer si existen diferencias entre los grupos propuestos.

4.20 Con fundamento en el planteamiento del problema 4.9, realiza una prueba de discriminante *LSD* de un factor con el propósito de establecer si existen diferencias entre los grupos propuestos.

4.21 Con base en el planteamiento del problema 4.10, realiza una prueba de discriminante *LSD* de un factor con el fin de establecer si hay diferencias entre los grupos propuestos.

SR **4.22** Una cooperativa agroindustrial ensaya con cinco tipos diferentes de aditivos proteínicos para la fermentación de sus quesos y procede a medir el pH. La empresa le encarga a un laboratorio en alimentos que les indique si existen diferencias entre los aditivos, por lo que los asesores llevan a cabo una prueba de ANOVA de un solo factor al 95% de confianza. En la tabla que sigue aparecen los resultados de las pruebas.

Tabla 4.34

Aditivo 1	Aditivo 2	Aditivo 3	Aditivo 4	Aditivo 5
4.01	4.07	4.20	4.20	4.14
4.04	4.12	4.17	4.15	4.12
4.02	4.16	4.21	4.19	4.14
4.04	4.15	4.19	4.17	4.14
4.03	4.08	4.20	4.14	4.11

4.23 Con base en el planteamiento del problema anterior, realiza un proceso de discriminación basado en la *R* de Duncan para establecer las diferencias entre los grupos.

SR **4.24** Una empresa fabricante de focos de vapor ensaya los nuevos modelos de focos de baja presión para uso comercial tratando de decidir entre dos modelos. De momento se realizan las pruebas de vida útil tal como se muestra en la siguiente tabla. Realiza una prueba de ANOVA de un factor de una muestra para comparar la vida útil entre los modelos de foco con un nivel de confianza del 95%.

Tabla 4.35

Modelo I Vida útil en horas	Modelo II Vida útil en horas
8 090	7 994
8 090	8 115
8 120	7 810
8 111	8 110
8 112	7 999
8 115	7 811
8 126	8 050
8 116	8 068
8 108	7 956
8 112	8 106

4.25 Con fundamento en los resultados del problema anterior, realiza un proceso de discriminación por *LSD*.

4.26 Un proceso de fabricación se reproduce en una planta por medio de tres máquinas al mismo tiempo, las 24 horas del día durante los siete días de la semana. El nivel de piezas producidas por día se muestra a continuación:

Tabla 4.36

	Máquina A	Máquina B	Máquina C
Lunes	400	395	398
Martes	385	393	397
Miércoles	410	390	399
Jueves	405	397	401
Viernes	406	399	401
Sábado	393	394	396
Domingo	400	395	398

Realiza una prueba de hipótesis estadística basada en ANOVA de dos factores con un error del 5%.

4.27 Con base en los resultados del problema anterior, realiza un proceso de discriminación por *LSD* para los factores *a* y *b*.

4.28 Con fundamento en los datos de la prueba de concreto, la empresa productora realiza pruebas sobre un nuevo aditivo acelerador para concreto con el cual se pretende estandarizar la resistencia en los concretos de 250 kg/cm^2 desarrollados con los tres tipos de cemento: puzolánico I, puzolánico II y puzolánico III. Lleva a cabo un análisis de hipótesis estadísticas basado en ANOVA de dos factores con varias muestras al 95% para determinar si existe o no diferencia en la resistencia de los concretos.

Tabla 4.37

	Resistencia en concreto con cemento Puzolánico I	Resistencia en concreto con cemento Puzolánico II	Resistencia en concreto con cemento Puzolánico III
Aditivo I	258	248	254
	263	252	256
	256	246	257
	256	248	255
Aditivo II	262	247	251
	262	246	253
	256	250	253
	258	250	254
Aditivo III	256	251	257
	254	250	256
	255	252	255
	256	255	254

Estadísticos e-Books & Papers

4.29 Con base en los resultados del problema anterior, haz un análisis por discriminantes *LSD* de dos factores incluyendo el análisis por interacción.

4.30 La empresa fabricante de focos no solo ensaya con el diseño de los nuevos modelos de focos de vapor de baja presión para uso comercial en cuanto a su vida útil, sino también con el uso del vapor de sodio y del de mercurio. Realiza una prueba de ANOVA de dos factores con varias muestras para determinar si los tipos de vapor y modelos generan diferencias en la vida útil de los focos. Considera un nivel de confianza del 95%.

Tabla 4.38

	Modelo I Vida útil en horas	Modelo II Vida útil en horas
Vapor de sodio	8 090	7 994
	8 090	8 115
	8 120	7 810
	8 111	8 110
	8 112	7 999
Vapor de mercurio	8 115	7 811
	8 126	8 050
	8 116	8 068
	8 108	7 956
	8 112	8 106

4.31 Con base en los resultados del problema anterior, lleva a cabo un análisis por discriminantes *LSD* de dos factores, incluye el análisis por interacción.

4.32 Una empresa fabricante de autos eléctricos utilitarios para usos especiales realiza pruebas sobre el rendimiento del banco de baterías de su nuevo modelo de auto ultraligero, el cual cuenta con tres versiones. Para el banco de baterías propone seis baterías de plomo ácido o seis de níquel-cadmio. La eficiencia eléctrica del banco en unidades porcentuales se muestra en la siguiente tabla. Realiza una prueba de ANOVA de dos factores al 95% de confianza para establecer si los tipos de banco y de auto generan diferencias en los rendimientos.

Tabla 4.39

	Auto tipo I	Auto tipo II	Auto tipo III
Plomo-ácido	82%	84%	78%
	78%	86%	79%
	79%	82%	80%
	74%	83%	82%
	80%	85%	77%
	75%	84%	80%
Níquel-cadmio	86%	92%	90%
	82%	88%	86%
	84%	85%	92%
	87%	90%	88%
	84%	86%	84%
	86%	83%	90%

4.33 Con base en los resultados del problema anterior, realiza un análisis por discriminantes *LSD* de dos factores incluyendo el análisis por interacción.

4.34 Un experimento en cuanto a la mejora del diseño de un helicóptero de aeromodelismo se fundamenta en el rendimiento de la distancia recorrida en razón del tipo de motor y del tipo de combustible empleado. Las distancias recorridas se muestran en la siguiente tabla.

Tabla 4.40

	Motor 1	Motor 2	Motor 3	Motor 4
Tipo 1	630	590	525	604
Tipo 2	680	5701	540	603
Tipo 3	670	610	534	595
Tipo 4	650	585	548	579
Tipo 5	640		554	594
Tipo 6	668			600

Realiza un análisis comparativo simple a efecto de determinar si existen o no diferencias entre los tipos de motor con respecto al tipo de combustible empleado.

4.35 Un experimento en cuanto a la mejora de un combustible para autos de tipo ecológico se estructuró con base en probar diferentes tipos de autos compactos. El experimento midió entre otros parámetros el rendimiento en kilómetros por litro. Estos se muestran en la siguiente tabla.

Tabla 4.41

Combustible	Auto 1	Auto 2	Auto 3	Auto 4	Auto 5	Auto 6
Tipo 1	12.70	13.10	14.10	12.40	14.20	13.14
Tipo 2	13.20	13.00	13.92	12.78	14.25	13.15
Tipo 3	12.70	13.05	13.89	12.80	14.30	13.20
Tipo 4	12.90	12.95	14.00	12.75	14.20	13.25
Tipo 5	12.82	12.92	13.90	12.82	14.16	13.17
Tipo 6	12.87	12.94	13.95	12.78	14.22	13.16
Tipo 7	12.92	12.97	13.98	12.80	14.22	13.17

Realiza una prueba de hipótesis basada en el ANOVA de dos factores de una muestra, al 95% de confianza, a efecto de determinar si existen o no diferencias entre los tipos de autos con respecto al tipo de combustible empleado.

4.36 Con fundamento en los resultados del problema anterior, realiza un análisis por discriminantes *LSD*. De manera adicional desarrolla un proceso de discriminación por coeficiente *LSD* para cada uno de los factores, así como para la interacción de los mismos.

4.37 Con base en el planteamiento del problema 4.35, realiza una prueba de hipótesis de un factor por tipo de auto de

Estadísticos e-Books & Papers

manera que se realice un análisis por discriminantes basado en la *R* de Duncan.

4.38 Con fundamento en el planteamiento del problema 35, procede a realizar una prueba de hipótesis de un factor por tipo de combustible de manera que se desarrolle un análisis por discriminantes basado en la *R* de Duncan.

PROBLEMA RETO

Un experimento en agronomía se estructuró con base en diferentes formas de tratar el suelo y sembrar una semilla mejorada de cierto tipo de oleaginosa. La cantidad sembrada dependió del clima y de la calidad del agua; las producciones en kilos se muestran en la tabla.

Tabla 4.42

	Trat 1	Trat 2	Trat 3	Trat 4	Trat 5
Cantidad 1	425	380	393	370	400
Cantidad 2	385	378	397	382	405
Cantidad 3	378	394	394	378	397
Cantidad 4	410	372	392	380	399
Cantidad 5	420	383	390	381	402
Cantidad 6	390	398	388	384	401
Cantidad 7	382	374	390	386	397
Cantidad 8	398	368	389	388	395
Cantidad 9	379	397	396	390	400
Cantidad 10	396	392	394	383	398

1

a) Haz un análisis comparativo simple para los tratamientos a efecto de determinar si existen diferencias entre los mismos.

b) Realiza un análisis comparativo simple para las cantidades de semilla a efecto de determinar si existen diferencias entre las mismas.

c) Desarrolla un proceso de discriminación basado en la *R* de Duncan para determinar si existen diferencias entre los tratamientos.

d) Realiza un proceso de discriminación basado en la *R* de Duncan para determinar si existen diferencias entre las cantidades.

e) Haz una prueba de hipótesis estadística por ANOVA de un factor para los tratamientos y compara estos resultados con los del inciso a). Argumenta, ¿existen diferencias entre los resultados?

f) Realiza una prueba de hipótesis estadística por ANOVA de un factor para las cantidades de semilla. Compara estos resultados con los del inciso b). Argumenta, ¿existen diferencias entre los resultados?

g) Desarrolla un análisis por discriminantes *LSD* para los tratamientos con el propósito de determinar si existen diferencias entre los mismos.

h) Con base en los resultados de los incisos c) y g), ¿existen diferencias entre los resultados de los discriminantes?

Estadísticos e-Books & Papers

i) Realiza un análisis por discriminantes *LSD* para las cantidades de semilla con el propósito de determinar si existen diferencias entre ellos.

j) Con base en los resultados de los incisos d) e i), ¿existen diferencias entre los resultados de los discriminantes?

k) Con fundamento en el planteamiento del problema reto, pon en práctica una prueba de hipótesis estadística basada en el ANOVA de dos factores con el propósito de determinar si existen diferencias tanto en los tratamientos, como en cantidades y la interacción entre ellos.

l) Realiza un análisis por discriminantes *LSD* de dos factores tanto para los tratamientos, como para las cantidades y la interacción entre ellos con el propósito de determinar si existen diferencias.

REFERENCIAS

Box, George E. P., William G. Hunter y J. Stuart Hunter (1993). *Estadística para investigadores-introducción al diseño de experimentos, análisis de datos y construcción de modelos.* España: Editorial Reverté, S.A.

Griesbrecht, F. G. y Marcia L. Gumperts (2004). *Planning, Construction and Statistical Analysis*. Estados Unidos: Wiley and Sons.

Kruehl, Robert O. (2001). *Diseño de experimentos* (2a. ed.). Estados Unidos: Thomson.

Montgomery, Douglas C. (2001). *Design and Analysis of Experiments* (5a. ed.). Estados Unidos: Wiley and Sons.

Vicente, Ma. Lina, Pedro Girón, Carmen Nieto y Teresa Pérez (2005). *Diseño de Experimentos–Soluciones SAS y SPSS*. México: Pearson-Prentice Hall.

DIRECCIONES ELECTRÓNICAS

Práctica 3. Diseño de experimentos con dos o más factores

[http://dm.udc.es/asignaturas/estadistica2/secprac_3.html]

Bases del análisis de la varianza

[http://www.hrc.es/bioest/Anova_2.html]

El análisis de la varianza (ANOVA)- 2. Estimación de componentes de varianza,
[http://argo.urv.es/quimio/general/anova2cast.pdf]

Tabla *t* de Student

[www.emp.uva.es/inf_acad/hermer/.../e2t_tabla_t_de_student.pdf]

Valores *F* de la distribución *F* de Fisher
[http://dcb.fi-c.unam.mx/profesores/irene/Notas/tablas/Fisher.pdf]

Tablas de rangos significativos de Duncan
[http://costaricalinda.com/Estadistica/duncan1.htm]

Fisher's *LSD* (Least Significant Difference)
[http://www.mzandee.net/~zandee/statistiek/syllabus/*LSD*.pdf]

An example of analysis of variance with two factors, compare means
[*www.seedtest.org/upload/cms/.user/ANOVA_2.pdf]*

579
Estadísticos e-Books & Papers